Freie Universität Berlin

The Gender Gap in Lifetime Earnings
The Role of Parenthood

Rick Glaubitz
Astrid Harnack-Eber
Miriam Wetter

School of Business & Economics
Discussion Paper

Economics

2022/3

The Gender Gap in Lifetime Earnings: The Role of Parenthood

Rick Glaubitz[*], Astrid Harnack-Eber[†], Miriam Wetter[‡]

Abstract

To obtain a more complete understanding of the persisting gender earnings gap in Germany, this paper investigates both the cross-sectional and biographical dimension of gender inequalities. Using an Oaxaca Blinder decomposition, we show that the gender gap in annual earnings is largely driven by women's lower work experience and intensive margin of labor supply. Based on a dynamic microsimulation model, we then estimate how gender differences accumulate over work lives to account for the biographical dimension of the gender gap. We observe an average gender lifetime earnings gap of 51.5 percent for birth cohorts 1964-1972. We show that this unadjusted gender lifetime earnings gap increases strongly with the number of children, ranging from 17.8 percent for childless women to 68.0 percent for women with three or more children. However, using a counterfactual analysis we find that the adjusted gender lifetime earnings gap of 10 percent differs only slightly by women's family background.

Keywords: Lifetime Earnings, Gender Inequality, Parenthood, Dynamic Microsimulation.

JEL-Classification: D31, J13, J16, J31

Acknowledgment: We thank Timm Bönke, Giacomo Corneo, Natalia Danzer, Richard Blundell, Carsten Schröder, Miriam Beblo, Daniel Waldenström and seminar participants at numerous venues and conferences for many valuable and helpful comments. Our thanks also go to the team of DIW Berlin for their guidance and data provision. Rick Glaubitz gratefully acknowledges financial support from the Hans-Böckler-Foundation.

[*]School of Business and Economics, Freie Universität Berlin; Guest researcher at DIW Berlin; email: `rick.glaubitz@fu-berlin.de`

[†]School of Business and Economics, Freie Universität Berlin;
email: `a.harnackeber@gmail.com`

[‡]School of Business and Economics, Freie Universität Berlin;
email: `miriam.wetter@fu-berlin.de`

1 Introduction

While most research on the gender pay gap has focused on differences in cross-sectional data, gender inequalities can add up over the life course as previous work experience, career pathways and earnings determine future labor market outcomes. Hence, a purely cross-sectional analysis cannot account for the biographical dimension of gender inequalities. But due to high data requirements, there is only scarce empirical evidence on gender lifetime earnings gaps. In addition, such studies are often limited by their use of administrative data and subsequent inability to account for women's family background, although women's household-related periods of labor market inactivity play an important role in gender differences over the work life. Therefore, it is crucial to account for women's family background and the role of motherhood when analyzing the gender lifetime earnings gap.

This study uses the Socio-economic Panel (SOEP) to shed light on the role of women's family backgrounds in gender differences, from both a cross-sectional and a lifetime perspective. In the cross-sectional analysis for birth cohorts 1940-1979, we find that the observed gender gap in annual earnings is more than twice as large as the hourly wage gap at almost every point in life after age 25. Using an Oaxaca Blinder decomposition, we show that the gaps can largely be explained by the extensive and intensive margins of labor. On average, women have less work experience and work less hours, which has a strong negative effect on women's earnings.

To further take advantage of the detailed socio-economic and family background information in the SOEP survey compared to administrative data sources, we use a dynamic microsimulation model to obtain full employment biographies, and subsequently lifetime earnings data. We can then estimate that women accumulate on average around 51.5 percent less lifetime earnings than men in terms of lifetime earnings up to the age 60. The unadjusted gender gap in lifetime earnings correlates largely with the number of children and ranges from 17.3 percent for childless women to 68.0 percent for women with three children or more.

To investigate which part of the observed gender gap in lifetime earnings can be associated with differences in the distribution of characteristics across gender and which part is due to differences in labor market returns to characteristics, we estimate women's counterfactual lifetime earnings. We find that around 80 percent of the observed lifetime earnings gap can be explained by different characteristics across men and women, leading to an adjusted gender lifetime earnings gap of 10 percent. Contrary to the unadjusted gap, motherhood does not play a crucial role for the adjusted gender lifetime earnings gap. The adjusted gender gaps in lifetime earnings for childless women and women with three and more children only differ by two percentage points.

Our paper is related to three different strands of literature. First, it contributes to the extensive literature on the gender gap in pay and its drivers.[1] Recent studies show that a large extent of the pay gap can be attributed to fewer hours worked and higher discontinuity of female employment biographies (e.g., Bertrand et al., 2010; Blau and Kahn, 2017). The persistence of this gender earnings inequality is mainly due to different effects of parenthood on men's and women's labor market behavior, and consequently their earnings (Waldfogel, 1998; Angelov et al., 2016; Kleven and Landais, 2017; Kleven et al., 2019). In line with previous studies (e.g., Goldin, 2014; Juhn and McCue, 2017; Gallen et al., 2019), we confirm that gender differences in annual earnings increase during the period of family formation, peak around age 40 and slowly decrease until retirement, leading to an inverse U-shape of the gender annual earnings gap over the work life.

[1] Past studies in this field focused on gender differences in human-capital accumulation and discrimination as the main drivers of gender inequalities in labor markets. Altonji and Blank (1999) give an overview of the early literature in this field.

Studies for Germany show that the cross-sectional earnings gap between mothers and non-mothers are largely driven by domestic work and childcare duties (e.g., Beblo and Wolf, 2002; Ejrnæs and Kunze, 2013). Strikingly, child penalties on women's pay are high in Germany compared to other countries. This is often attributed to longer maternal leave entitlement and a higher rate of part-time work for women in Germany (e.g., Harkness and Waldfogel, 2003; Gangl and Ziefle, 2009; Kleven et al., 2019). We confirm that the gender gap in annual earnings increases with the number of children and that motherhood plays a crucial role in women's extensive and intensive labor margins.

The literature on the gender pay gap and its evolution has primarily focused on cross-sectional hourly wages, annual earnings or earnings over a short period of time. Hence, the empirical evidence on how gender inequalities add up or equalize over the lifecycle is scarce – this is also due to high data requirements. Further, such studies have often only focused on employed individuals, not accounting for periods of inactivity or unemployment which are in fact often the results of household-related labor supply changes, especially for women (Kleven et al., 2019). Using administrative data, Guvenen et al. (2014) show that the fraction of women among lifetime top earners is significantly lower than that of men for birth cohorts 1956 to 1958. On average, lifetime top earners in the U.S. tend to be individuals who experience high earnings growth over the first half of their lifecycle – the period when the gender gap increases the most, likely due to family-related reasons. In a later study, Guvenen et al. (2017) provide evidence that the large gender lifetime earnings gap is narrowing over time, with women's median lifetime earnings increasing while men's median lifetime earnings decreasing for younger birth cohorts.

Using administrative data from the German Pension Register (VSKT), Bönke et al. (2015) find evidence that intragenerational lifetime earnings inequality for West German men born between 1935 and 1969 has increased, largely due to losses in the bottom of the lifetime earnings distribution. They also supplement their work with additional results on West German women. However due to data restrictions, their data only includes women with stable employment biographies. Therefore, the VSKT data is not representative for most women mainly due to the high rate of inactivity amongst women and cannot be used for estimating the gender lifetime earnings gap in Germany.

Closest to our paper is the study by Boll et al. (2017) analyzing the gender lifetime earnings gap in Germany. Using the administrative Sample of Integrated Labour Market Biographies (SIAB), they estimate an unadjusted gender lifetime earnings gap of 46 percent for West German birth cohorts 1950 to 1964. They show that the gender gap widens significantly during the age of family formation and that gender differences in work experience and hours worked explains around two-thirds of this overall gender lifetime earnings gap. However, SIAB data does not offer any information about individuals' family background. Hence, to the best of our knowledge, our study is the first to control for the influence of motherhood when estimating the adjusted gender lifetime earnings gap in Germany. While we find a similar unadjusted gender lifetime earnings gap of 51.5 percent, our counterfactual analysis can explain up to 80 percent of the observed gap in lifetime earnings by additionally controlling for individuals' family backgrounds. This underlies the importance of including the family background to understand women's employment biographies. To be able to investigate lifetime earnings and the role of motherhood together, we apply a dynamic microsimulation model to SOEP survey data to obtain full lifetime earnings data up to age 60. Combining simulation models with survey data is a well-established method to deal with missing observations and panel attrition, which often impede using survey data to conduct research on lifetime earnings or long-term analyses (e.g., Brown et al., 2009; Coronado et al., 2011).

Our simulation approach is closest to studies implementing a regression-based simulation approach predicting the transition probabilities of individuals or households moving from one state to another between two different points in time (e.g., Heien et al., 2007; Geyer and Steiner, 2014). In contrast to studies using a splicing approach (e.g., Grabka and Goebel, 2017; Westermeier et al., 2012) where sequences of existing biographies are stitched together to construct full lifecycle data, our approach typically "ages " the data year by year (Li and O'Donoghue, 2013). The simulation approach by Levell and Shaw (2015) is the closest to ours.[2]

The next section introduces our dataset and starts by analyzing cross-sectional gender differences in hourly wages and annual earnings over the work life by using an Oaxaca Blinder decomposition. Section 3.1 describes our microsimulation approach to obtain full work biographies, while Section 3.2 estimates the gender lifetime earnings gap and further investigates the role of motherhood. Section 4 concludes.

2 Cross-Sectional Analysis

The cross-sectional analysis allows us to explore how gender gaps in hourly wages and annual earnings develop with increasing age and to investigate if short-term differences already follow certain patterns across gender. This first step is crucial to subsequently better understand how gender inequalities in labor market characteristics and earnings add up or equalize over the entire work life.

2.1 Data and Methodology

Our study is based on the 35th wave (1984-2018) of the German Socio-Economic Panel (SOEP). The SOEP is a representative annual panel survey questioning about 30,000 individuals across 15,000 households since 1984. In contrast to administrative data, the SOEP includes a rich set of socio-economic variables, detailed labor market

[2] Other studies close to ours are the ones by Bonin et al. (2015) and Hänisch and Klos (2016) which also simulate employment biographies using SOEP data.

information and household background including information on the partner and children.[3]

We restrict our cross-sectional analysis to birth cohorts 1940 to 1979. These are the same birth cohorts used for the underlying regressions of our microsimulation model in Chapter 3. We observe these cohorts at least once between the ages of 38 and 44 in the SOEP. This age restriction is crucial as it is the age frame when individuals' cross-sectional earnings show the highest correlation with lifetime earnings (Björklund, 1993; Bönke et al., 2015) and is therefore needed to successfully simulate life cycle profiles in Section 3. Further, we focus on West German individuals since those born in East Germany were only included in the SOEP after German reunification in 1990. The poor comparability of the Federal Republic of Germany and the German Democratic Republic with respect to labor market institutions and economic systems does not allow us to simulate missing information for East Germans before 1990.

Section 2 focuses on the evolution of cross-sectional hourly wages and annual earnings with increasing age over the work life. This approach sheds light on two main components of the gender gap in lifetime earnings; the gender gap in hourly wages shows the differences in the compensation between women and men for one hour of their work, while the gap in annual earnings reveals dissimilarities driven by the variation in working hours.

We use an Oaxaca Blinder decomposition (see Oaxaca, 1973; Blinder, 1973) to investigate how much of the difference in the observed gender gap is driven by different characteristics between men and women and how much can be attributed to different returns to characteristics within the labor market.[4] Using this decomposition approach the gender gap G in the labor market outcome variable L (here:

[3]See Goebel et al. (2019) for a detailed overview about the SOEP.

[4]A more detailed description of this methodological approach can be found in Section A.1, Appendix A.

logarithmic hourly wage and logarithmic annual earnings) is defined as:

$$G_x = E(L_{mx}) - E(L_{fx}) \tag{1}$$

Therefore, G is the gender differential between the means of outcome L for men (m) and women (f) at age x. We can then divide the gender gap into two parts. First, the *endowment part*, which is the component of the gender gap which is due to differences in the distribution of characteristics between men and women. And second, the *coefficient part*, which accounts for differences in returns to characteristics. Hence, the coefficient part shows the gender driven difference of the labor market's willingness to pay for the same characteristics obtained by either men or women. However, note that the coefficient part may also include gender differences that remain unexplained in our model due to data and model restrictions. We run the following regression model separately by sex (s) and age (x) for the labor outcome L [5]:

$$L_{s,i,x} = \alpha_{s,i,x} + \beta_{s,i,x} Z_{s,i,x} + \epsilon_{s,i,x}, \quad E(\epsilon_{s,x}) = 0, \quad s \in \{F, M\}, x \in [20, 60] \tag{2}$$

where Z is a vector of control variables including work experience measured as number of working years, full-time or part-time work, work sector, highest education level, marital status and number of kids. In addition, we control for cohort and time effects.[6]

[5] For comparability, we only control for variables that we can also use in our analysis of the lifetime gender gap in Chapter 3.

[6] Our pooled sample includes birth cohorts 1940 to 1979. Therefore, we include cohort dummies into our estimation model. We do not find any consistent cohort effects in our analysis. Hence, results are stable for all cohorts.

Table 1: Descriptive Statistics - Means by Age

Men	Age 20	25	30	35	40	45	50	55	60
Annual earnings	15748 (10972)	27727 (13307)	37925 (18571)	45217 (24095)	51615 (31182)	54204 (38951)	54747 (35380)	53969 (33505)	51535 (50496)
Hourly wage	9.37 (7.72)	15.13 (18.11)	18.12 (20.76)	20.72 (16.32)	23.06 (16.25)	23.95 (17.91)	24.24 (14.47)	25.83 (31.96)	26.13 (28.33)
Hours worked per week	34.55 (13.42)	38.29 (14.48)	42.81 (12.47)	43.49 (11.38)	44.39 (11.01)	44.10 (10.61)	43.50 (11.17)	42.65 (12.13)	39.34 (14.38)
Years in full-time work	1.20 (1.28)	4.75 (2.60)	8.54 (3.77)	12.97 (4.37)	17.71 (4.85)	22.58 (5.31)	27.43 (5.71)	32.69 (5.81)	37.32 (5.74)
Years in part-time work	0.14 (0.47)	0.33 (0.98)	0.55 (1.56)	0.56 (1.64)	0.61 (1.93)	0.65 (2.10)	0.75 (2.44)	0.68 (2.46)	1.09 (2.90)
Years in unemployment	0.13 (0.38)	0.31 (0.70)	0.39 (0.98)	0.43 (1.21)	0.45 (1.37)	0.47 (1.64)	0.51 (1.85)	0.49 (1.78)	0.45 (1.65)
Years of education	8.97 (3.86)	10.61 (3.03)	11.84 (3.25)	12.44 (3.17)	12.62 (3.03)	12.65 (2.96)	12.67 (2.92)	12.57 (2.84)	12.73 (2.92)
Women	**20**	**25**	**30**	**35**	**40**	**45**	**50**	**55**	**60**
Annual earnings	12773 (8683)	21115 (12332)	22975 (16720)	21925 (19512)	22944 (18626)	24975 (20497)	26705 (21713)	26475 (25559)	24659 (21236)
Hourly wage	7.97 (6.58)	12.87 (9.59)	15.19 (12.19)	15.63 (13.18)	16.23 (12.02)	16.23 (10.88)	16.82 (12.49)	16.54 (13.18)	17.48 (14.99)
Hours worked per week	31.75 (13.02)	32.28 (14.34)	29.91 (15.60)	26.68 (15.26)	27.38 (14.29)	28.87 (13.99)	30.09 (13.98)	29.42 (13.64)	26.97 (14.49)
Years in full-time work	1.20 (1.23)	4.36 (2.71)	6.73 (4.12)	8.04 (5.23)	9.63 (6.46)	11.61 (7.99)	14.00 (9.69)	16.70 (11.75)	19.65 (13.99)
Years in part-time work	0.21 (0.55)	0.82 (1.61)	1.91 (2.60)	3.81 (3.76)	5.69 (4.87)	7.60 (6.16)	9.45 (7.75)	11.66 (9.77)	13.32 (11.79)
Years in unemployment	0.17 (0.40)	0.28 (0.73)	0.40 (0.95)	0.50 (1.19)	0.56 (1.52)	0.58 (1.60)	0.64 (1.77)	0.71 (1.99)	0.57 (1.86)
Years of education	9.17 (3.91)	11.17 (2.98)	12.07 (3.31)	12.42 (3.00)	12.48 (2.89)	12.39 (2.91)	12.34 (2.78)	12.11 (2.59)	12.02 (2.71)

Note: Only employed individuals with hourly wages and annual earnings greater than zero were included. Cohorts 1940-1979, weighted sample. Annual earnings and hourly wages in 2015 prices (Euro). Standard errors in parentheses.
Source: Own calculations based on SOEP v.35 (1984-2018).

2.2 Hourly Wage

Overall, employed men have significantly higher hourly wages than employed women (see Table 1). At the beginning of their work life at age 20, men earn on average 9.37 Euro per hour while women's average wage is only 7.97 Euro per hour in 2015 prices. In line with results found by Federal Statistical Office (2017), the average hourly wages of men in our sample then almost triples over the work life to 26.13 Euro per hour at age 60. In contrast, women's hourly wages only increase to 17.48 Euro, already showing significant gender differences in wage growth over the work life.

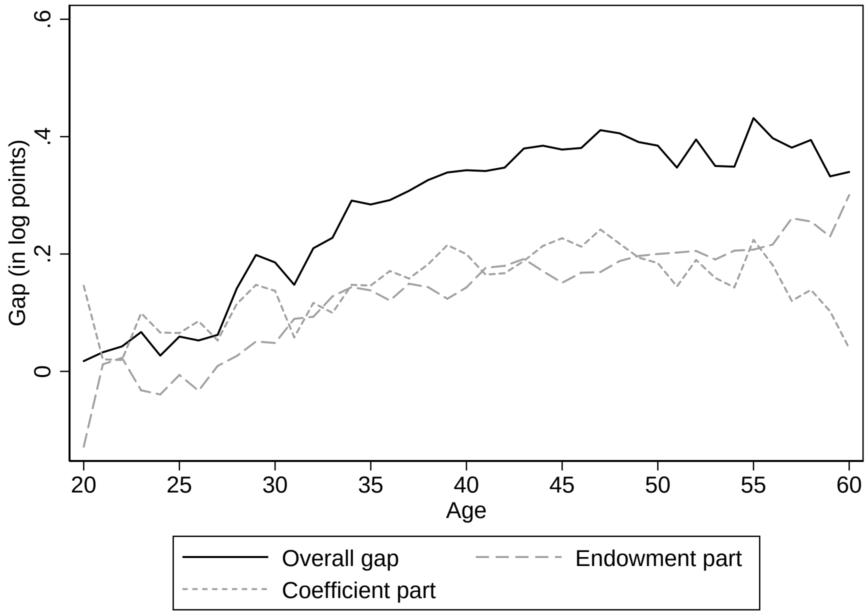
The solid line in Figure 1 shows the evolution of the gender gap in hourly wages in log points from age 20 to 60. Notably, the gender gap remains stable over the early years of work life. At age 25, men's hourly wages are only 0.059 log points higher than women's and the difference is still insignificant (see also Table 2). However, during the time of family formation and childcare, this gap drastically widens up to a highly significant difference of 0.378 log points at age 45.[7] Afterwards, the growth of the gender gap in hourly wages slows down and remains relatively stable with a peak at age 55. This finding is consistent for all cohorts (see Figure A.2, Appendix A). In line with our findings, previous studies also documented a widening of the gender wage gap over the lifecycle (e.g., Anderson et al., 2002; Angelov et al., 2016; Tyrowicz et al., 2018).[8]

The results of the Oaxaca Blinder decomposition are displayed by the grey lines in Figure 1 and also in Table 2. Visibly, the widening of the gender gap in hourly wages over the work life is driven by the increase in the endowment part, while

[7] A gender gap of 0.059 log points corresponds to a wage differential of $(e^{0.059} - 1) * 100 = 6.08$ percent, while a gap of 0.378 log points corresponds to a wage differential of $(e^{0.378} - 1) * 100 = 45.94$ percent.

[8] To show that our findings are in line with the broad literature on the cross-sectional gender gap in hourly wages, Figure A.1, Appendix A shows the gender gap in hourly wages for our sample by survey year.

Figure 1: Gender Gap in Hourly Wages



Notes: Only employed individuals are considered. Cohorts 1940-1979, weighted sample.

Source: Own calculations based on SOEP v.35 (1984-2018).

the coefficient part of the gender gap shapes its overall trend. At younger ages, the different distribution of characteristics does not play a role yet. Therefore, at the beginning of work life all wage differences between men and women are due to different returns to labor market characteristics. Main differences in characteristics such as work experience or family background widen only later in life; after age 25, the high and significant coefficients for work experience in Table 2 show that the increase of the endowment part is mainly driven by women's lower gain of work experience with increasing age. On average, women stay at home for childrearing more often and for longer periods than men do. Consequently, after starting a family, they participate in the labor market to a lesser extent or not at all for many years. By the age of 60, men have accumulated on average 37.32 years of full-time and 1.09 years of part-time work experience, whereas women have accumulated on average only 19.65 years of full-time and 13.32 years of part-time work experience (see Table 1).

The Oaxaca Blinder decomposition shows that these large differences in work experience are crucial to explaining the gender gap in hourly wages. By the end of work life, differences in work experience account for 0.309 log points of the overall gender wage gap of 0.340 log points. Hence, around 90 percent of the overall gender gap of 40.5 percent in hourly wages can be explained by differences in work experience. Since work experience is different from characteristics that are more stable later in life like education or number of children, it has a very dynamic influence on the overall earnings potentials of women. Therefore, it is crucial to shed more light on its role while analyzing the gender gap in annual and lifetime earnings.

Regarding the influence of the distribution of education, it is important to note that it is shaped by cohort effects rather than effects over the life cycle. For younger generations, which we observe at younger ages, our data shows that women have the tendency to have on average a higher or equal degree of education as men. Contrarily, for older generations which we can only observe at older ages, men had a higher level of education than women on average. Therefore, for older ages, education has a positive explanatory power as women used to be less educated than men. For younger cohorts, women's educational advantage decreases the endowment part, and thus the overall gender gap in hourly wages at an early age.

In contrast to the stable growth of the endowment part, the evolution of the coefficient part follows a slight inverse U-shape. At age 20, the gender gap cannot be explained through differences of characteristics across genders, but the coefficient part amounts to 0.126 log points. This means that even if women and men had the same labor market characteristics, men's wages would be 0.126 log points (13.4 percent) higher than women's wages at this age. The coefficient part of the gender gap then peaks at 0.247 log points (28.0 percent) at age 45 and then declines again to a difference of 0.042 log points (4.3 percent) just before retirement.[9] In contrast to the endowment part, none of the variable groups have a constant significant influence on the overall gender gap, including the constant itself.[10] Therefore, not one individual effect dominates the coefficient part of the overall gender gap, but the coefficient part is instead a combination of many individual influences including those not controlled for in this regression model.

In summary, the gender gap in hourly wages is determined by two factors: first, women have in sum less favorable labor market characteristics compared to men, and second, even if they had the same characteristics, the labor market rewards women worse than men. The influence of differences in characteristics grows significantly with age, mainly through increasing differences in accumulated work experience across gender. Of the observed gender gap of 40.5 percent (0.340 log points) at age 60, different characteristics account for 87 percent (0.297 log points). This leads to an adjusted gender gap in hourly wages of 4.3 percent.

However, when interpreting these results, we need to keep in mind that our model does not control for endogenous choice. We do not control for the fact whether women choose to leave the labor market for an extended period or work part-time or if they are forced to do so. The same holds true for occupational choices since women more often work in lower paid occupations. Nevertheless, we show that one hour of women's work is less rewarded than an hour of men‘s work - whether it is by occupational choice or by discrimination on the labor market.

[9]Table A.1 and Table A.2 in Appendix A display the separate regression results for men and women which provide the basis for the difference in coefficients displayed in the Oaxaca Blinder regression.

[10]The constant of the coefficient part also includes the effects of gender differences in unobserved predictors Jann (2008), e.g., different occupational choices or differences in employers.

Table 2: Oaxaca Blinder Decomposition of Hourly Wage Gender Gap

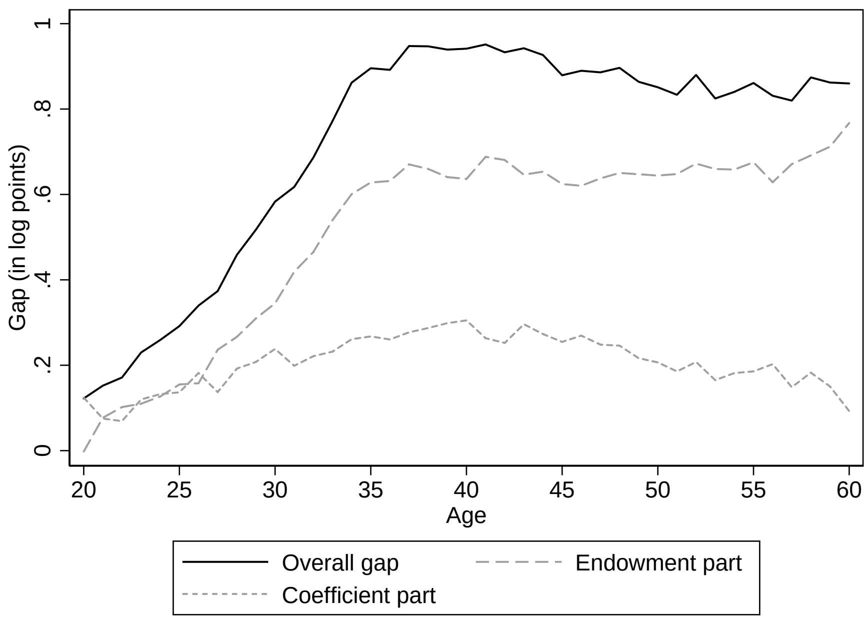
	(1) Age 20	(2) Age 25	(3) Age 30	(4) Age 35	(5) Age 40	(6) Age 45	(7) Age 50	(8) Age 55	(9) Age 60
Overall									
Men	1.963***	2.563***	2.771***	2.912***	2.980***	3.008***	3.019***	3.054***	3.003***
	(0.040)	(0.021)	(0.015)	(0.012)	(0.012)	(0.013)	(0.016)	(0.022)	(0.026)
Women	1.945***	2.503***	2.586***	2.628***	2.637***	2.630***	2.634***	2.622***	2.663***
	(0.033)	(0.024)	(0.019)	(0.017)	(0.016)	(0.017)	(0.017)	(0.020)	(0.031)
Difference	0.018	0.059	0.186***	0.284***	0.343***	0.378***	0.385***	0.432***	0.340***
	(0.052)	(0.031)	(0.024)	(0.021)	(0.020)	(0.021)	(0.024)	(0.030)	(0.041)
Endowment	-0.108**	-0.031	0.033*	0.107***	0.107***	0.131***	0.196***	0.200***	0.297***
	(0.040)	(0.018)	(0.015)	(0.018)	(0.020)	(0.022)	(0.026)	(0.028)	(0.043)
Coefficient	0.126*	0.091**	0.152***	0.177***	0.235***	0.247***	0.189***	0.231***	0.042
	(0.050)	(0.033)	(0.028)	(0.026)	(0.029)	(0.028)	(0.036)	(0.040)	(0.060)
Endowment									
Children	0.003	-0.000	0.001	-0.002	-0.004	-0.009*	-0.015*	-0.014	-0.013
	(0.005)	(0.003)	(0.001)	(0.002)	(0.003)	(0.004)	(0.007)	(0.011)	(0.013)
Married	0.001	-0.001	0.004	0.002	0.003	0.005*	0.000	-0.002	0.010
	(0.004)	(0.004)	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)	(0.003)	(0.008)
Experience	-0.068*	0.024*	0.084***	0.167***	0.207***	0.224***	0.228***	0.264***	0.309***
	(0.030)	(0.011)	(0.013)	(0.019)	(0.022)	(0.028)	(0.026)	(0.032)	(0.049)
Part time	0.002	-0.041**	-0.027	-0.067***	-0.096***	-0.066***	-0.026	-0.056*	-0.030
	(0.008)	(0.013)	(0.015)	(0.019)	(0.019)	(0.019)	(0.019)	(0.024)	(0.024)
Education	-0.008	-0.019**	-0.009	0.021**	0.019**	0.022***	0.031***	0.030***	0.045***
	(0.007)	(0.006)	(0.006)	(0.007)	(0.006)	(0.006)	(0.007)	(0.008)	(0.011)
Cohort	-0.001	-0.002	-0.000	0.001	0.003	0.001	0.002	0.003	-0.001
	(0.005)	(0.003)	(0.002)	(0.002)	(0.002)	(0.001)	(0.001)	(0.002)	(0.002)
Sector	-0.037	0.009	-0.019*	-0.017**	-0.025***	-0.045***	-0.024***	-0.024*	-0.022*
	(0.029)	(0.011)	(0.008)	(0.006)	(0.007)	(0.007)	(0.007)	(0.010)	(0.011)
Coefficient									
Children	0.003	0.030	0.095***	0.063	0.002	0.007	-0.075	-0.048	0.080
	(0.005)	(0.020)	(0.025)	(0.034)	(0.038)	(0.046)	(0.044)	(0.053)	(0.088)
Married	-0.006	0.021	0.041	0.052	0.090*	-0.008	0.100*	0.047	0.020
	(0.010)	(0.027)	(0.031)	(0.034)	(0.038)	(0.041)	(0.045)	(0.056)	(0.074)
Experience	0.243***	0.207	0.067	-0.094	-0.058	0.005	-0.150	0.196	-0.973
	(0.063)	(0.129)	(0.128)	(0.127)	(0.112)	(0.170)	(0.226)	(0.280)	(0.995)
Part time	0.013	0.031	0.008	-0.021	0.032	0.032	0.030	0.145***	0.036
	(0.027)	(0.037)	(0.023)	(0.019)	(0.021)	(0.021)	(0.025)	(0.036)	(0.054)
Education	0.329	-0.144	-0.210	-0.105	-0.564*	0.183	0.568	0.131	0.722
	(0.195)	(0.174)	(0.141)	(0.230)	(0.256)	(0.237)	(0.295)	(0.840)	(0.648)
Cohort	0.083	0.023	0.002	-0.007	0.091	-0.019	0.030	-0.030	0.002
	(0.044)	(0.085)	(0.036)	(0.125)	(0.049)	(0.034)	(0.041)	(0.038)	(0.040)
Sector	0.222	-0.445**	-0.436**	-0.053	-0.066	-0.139	-0.037	-0.085	-0.153
	(0.193)	(0.139)	(0.160)	(0.116)	(0.147)	(0.142)	(0.124)	(0.188)	(0.192)
Constant	-0.762**	0.368	0.587*	0.343	0.708*	0.186	-0.276	-0.125	0.308
	(0.276)	(0.283)	(0.255)	(0.308)	(0.322)	(0.317)	(0.383)	(0.887)	-1.090
N	765	1782	3053	4323	5356	5592	4304	2866	1758

Note: Standard errors in parentheses; The stars refer to the following significance level: $*p < 0.1$, $**p < 0.05$, $***p < 0.01$. The different drivers are summarized as followed: "Children": Number of children; "Married": Dummy variable on marital status, "Experience": Total years of working full time, part time or being inactive (also squared); "Part time": Dummy variable indicating full time or part time work; "Education": Dummy variables in highest level of educational attainment, "Sector": Occupational sector; "Cohort": Cohort dummies. Cohorts 1940-1979, weighted sample.
Source: Own calculations based on SOEP v.35 (1984-2018).

2.3 Annual Earnings

In addition to earning less per hour, women also work on average fewer hours than men do. Therefore, the gender gap in annual earnings might be even wider than the gap in hourly wages due to gender differences in the intensive margin of work.

Figure 2: Gender Gap in Annual Earnings



Notes: Only employed individuals are considered. Does not include values of zero annual earnings. Cohorts 1940-1979, weighted sample.

Source: Own calculations based on SOEP v.35 (1984-2018).

Figure 2 shows the overall gender gap in annual earnings, the part of the gap due to different characteristic across gender (endowment part) and the part of the gender gap due to differences in coefficients (coefficient part). Visibly, the gender gap in annual earnings is significantly higher than the gender gap in hourly wages. At the peak of the gap at age 40 (0.829 log points corresponding to 129.1 percent), men's average annual earnings are more than twice as high than women's. Similar to the gender gap in hourly wages, the gender gap in annual earnings increases rapidly until age 35 and remains on a constant high level during the years of child rearing. Afterwards, it only declines slightly until retirement. This finding is in line with earlier studies for the U.S. providing evidence for a similar course of the cross-sectional gender gap in annual earnings over the work life (Goldin, 2014; Juhn and McCue, 2017).

When decomposing the overall gender gap in annual earnings, we find that the larger gap is driven by the significantly higher endowment part. While the gender gap due to differences in coefficients is only slightly higher than in the model for hourly wages, the endowment part has more than tripled.[11] This result underlines the importance of differences in the intensive labor margin across gender.

Table 3 shows that the endowment part of the gender gap in annual earnings is also driven by the lesser work experience women have accumulated over their life cycle. Moreover, the lower number of hours worked by women per year at all ages influences the gender gap to an even greater extent. These findings are in line with previous studies (e.g., Bertrand et al., 2010; Gallen et al., 2019).

At the peak of family formation and child rearing around age 35, women's annual earnings are on average 0.327 log points lower than men's due to their fewer number of work hours.[12] In addition, women's earnings are on average 0.203 log points lower than men's due to the lesser work experience they have accumulated up to this age. This means that more than half of the overall gap can be explained by the distribution of working hours and a quarter can be explained by the different

[11] Please note that since this chapter focuses on the intensive margin of work, we now include the total hours worked per year for this model in contrast to the binary variable (part-time/full-time) used when we were analyzing the gender gap in hourly wages. Consequently, this leads to an even more significant endowment part for the analysis of annual earnings as the total number of work hours is a key driver in the earnings difference across gender.

[12] It is crucial to note that our model cannot control for endogenous choice. Hence, we do not differentiate whether women choose to work fewer hours or if they have trouble finding adequate employment after exiting the labor market for maternal leave. For example, Kleven et al. (2019) show for Germany that there are significant negative effects of childbirth on women's labor market participation and annual earnings, both in the short- and long-run.

distribution of work experience. The effect of work experience steadily increases over the life cycle and peaks at age 60 with 0.351 log points. In contrast, differences in the level of education or family background play a smaller role.

The coefficient part of the gender gap in annual earnings is positive throughout the life cycle. This means that besides the worse characteristics women also face less beneficial coefficients in their wage regression (see Table 3, and Tables A.3 and A.4 in Appendix A). This is especially pronounced in the years of child rearing. There are two potential explanations: First, employers could fear a higher risk of work absence by women due to pregnancy and child rearing, and therefore already include the higher risk of absence in the paid wages of women (e.g., Correll et al., 2007). Second, women might opt for less financially rewarding positions in return for higher work flexibility for the time after childbirth (e.g., Goldin, 2014). In summary, the evolution of the gender gap in annual earnings over the life cycle is both driven by differences in labor supply of men and women and by the gender gap in hourly wages.

Table 3: Oaxaca Blinder Decomposition of the Annual Earnings Gender Gap

	(1) Age 20	(2) Age 25	(3) Age 30	(4) Age 35	(5) Age 40	(6) Age 45	(7) Age 50	(8) Age 55	(9) Age 60
Overall									
Men	9.462***	10.155***	10.460***	10.623***	10.695***	10.713***	10.717***	10.693***	10.542***
	(0.041)	(0.024)	(0.017)	(0.012)	(0.013)	(0.015)	(0.017)	(0.022)	(0.030)
Women	9.424***	9.950***	9.923***	9.854***	9.867***	9.915***	9.904***	9.868***	9.775***
	(0.039)	(0.026)	(0.026)	(0.024)	(0.021)	(0.021)	(0.024)	(0.028)	(0.040)
Difference	0.038	0.205***	0.537***	0.769***	0.829***	0.797***	0.812***	0.825***	0.766***
	(0.057)	(0.036)	(0.031)	(0.027)	(0.024)	(0.025)	(0.030)	(0.036)	(0.050)
Endowment	-0.081	0.102***	0.318***	0.538***	0.555***	0.529***	0.609***	0.657***	0.747***
	(0.046)	(0.026)	(0.028)	(0.025)	(0.025)	(0.028)	(0.034)	(0.036)	(0.050)
Coefficient	0.119*	0.103**	0.219***	0.231***	0.274***	0.269***	0.203***	0.168***	0.019
	(0.046)	(0.036)	(0.036)	(0.028)	(0.031)	(0.033)	(0.042)	(0.042)	(0.053)
Endowment									
Children	-0.000	0.001	0.001	-0.001	-0.003	-0.010*	-0.018*	-0.020	-0.012
	(0.005)	(0.003)	(0.002)	(0.002)	(0.003)	(0.005)	(0.007)	(0.011)	(0.012)
Married	0.001	0.001	0.005	0.001	0.001	0.001	-0.002	-0.004	0.003
	(0.004)	(0.003)	(0.003)	(0.001)	(0.002)	(0.003)	(0.002)	(0.003)	(0.008)
Experience	-0.058	0.033*	0.129***	0.203***	0.244***	0.273***	0.306***	0.334***	0.351***
	(0.030)	(0.014)	(0.015)	(0.018)	(0.022)	(0.026)	(0.028)	(0.033)	(0.040)
Hours worked	0.023	0.082***	0.214***	0.327***	0.313***	0.282***	0.310***	0.331***	0.371***
	(0.018)	(0.015)	(0.024)	(0.022)	(0.021)	(0.023)	(0.025)	(0.029)	(0.034)
Education	-0.003	-0.021**	-0.011	0.023**	0.020**	0.023***	0.032***	0.033***	0.048***
	(0.006)	(0.007)	(0.007)	(0.007)	(0.006)	(0.007)	(0.007)	(0.008)	(0.012)
Cohort	-0.001	-0.002	0.000	0.001	0.002	0.000	0.002	0.003	-0.002
	(0.005)	(0.003)	(0.002)	(0.002)	(0.002)	(0.001)	(0.001)	(0.003)	(0.002)
Sector	-0.043	0.008	-0.019*	-0.016*	-0.021**	-0.040***	-0.021**	-0.020	-0.012
	(0.026)	(0.011)	(0.009)	(0.006)	(0.007)	(0.008)	(0.008)	(0.010)	(0.011)
Coefficient									
Children	0.000	0.037*	0.155***	0.069	-0.005	0.025	-0.093	-0.074	0.069
	(0.004)	(0.015)	(0.028)	(0.037)	(0.040)	(0.049)	(0.048)	(0.057)	(0.081)
Married	-0.004	0.033	0.040	0.049	0.092*	-0.012	0.113*	0.075	0.060
	(0.010)	(0.025)	(0.033)	(0.036)	(0.040)	(0.048)	(0.048)	(0.055)	(0.074)
Experience	0.224***	0.146	0.045	-0.139	-0.159	-0.150	-0.169	0.378	-0.964
	(0.063)	(0.139)	(0.146)	(0.156)	(0.121)	(0.177)	(0.269)	(0.423)	(0.966)
Hours worked	-0.348	-0.260	-0.625***	-0.811***	-0.801***	-0.783***	-0.891***	-1.152***	-0.621***
	(0.222)	(0.184)	(0.160)	(0.105)	(0.100)	(0.143)	(0.144)	(0.145)	(0.151)
Education	0.356*	-0.082	-0.240	-0.165	-0.595*	0.229	0.720*	0.932	0.809
	(0.173)	(0.182)	(0.151)	(0.245)	(0.263)	(0.310)	(0.332)	(0.700)	(0.699)
Cohort	0.067	0.002	-0.038	0.028	0.134**	-0.005	0.039	-0.030	0.003
	(0.044)	(0.051)	(0.039)	(0.124)	(0.046)	(0.036)	(0.044)	(0.040)	(0.040)
Sector	0.166	-0.295*	-0.396*	-0.126	-0.055	-0.091	-0.008	-0.120	-0.098
	(0.180)	(0.144)	(0.166)	(0.130)	(0.159)	(0.150)	(0.142)	(0.213)	(0.217)
Constant	-0.344	0.521	1.278***	1.326***	1.663***	1.056*	0.492	0.159	0.759
	(0.330)	(0.304)	(0.310)	(0.358)	(0.343)	(0.413)	(0.465)	(0.834)	-1.112
N	765	1782	3053	4323	5356	5592	4304	2866	1758

Note: Standard errors in parentheses; The stars refer to the following significance level: $*p < 0.1$, $**p < 0.05$, $***p < 0.01$. The different drivers are summarized as followed: "Children": Number of children; "Married": Dummy variable on marital status, "Experience": Total years of working full time, part time or being inactive (also squared); "Hours worked": Hours worked per year; "Education": Dummy variables in highest level of educational attainment, "Sector": Occupational sector; "Cohort": Cohort dummies. Cohorts 1940-1979, weighted sample.
Source: Own calculations based on SOEP v.35 (1984-2018).

3 Microsimulation and Lifetime Analysis

The cross-sectional analysis gave an informative overview on the short-term gender gap and its development with increasing age. In this chapter, we investigate how advantages or disadvantages might accumulate or balance out over the complete work life by looking at lifetime earnings. This allows us to shed light on the biographical dimension of the gender earnings gap.

3.1 Data and Methodology

We continue to use the SOEP as it offers long-term panel data with detailed labor market and family background information, which administrative data cannot offer. However, the SOEP suffers from panel mortality. Only around 10 percent of the participants have been observed for at least 20 years or more, with an average participation period is 9.36 years (see Figure B.1, Appendix B). To investigate lifetime earnings for a larger sample, we implement a dynamic microsimulation approach by regression to fill in the missing data of non-observed years during an individual's work live. This approach yields complete data for the whole period of our analysis while still benefitting from the rich set of socioeconomic characteristics and family information in the SOEP.

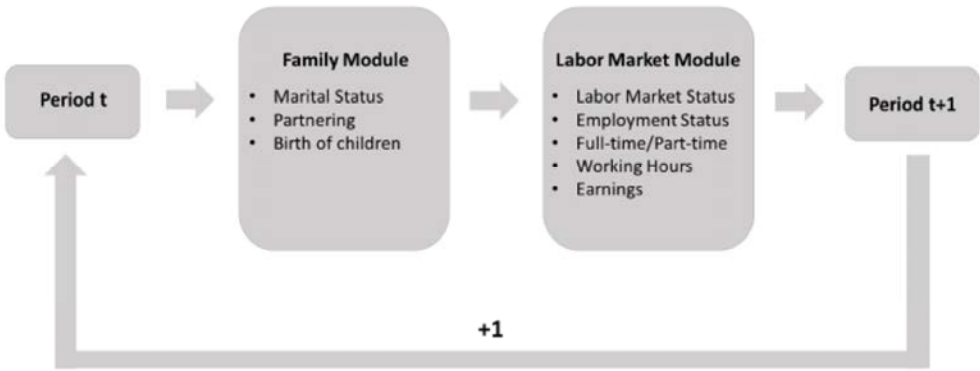
To implement our dynamic microsimulation model successfully, we need to add the following restrictions to our cross-sectional sample: First, our lifetime earnings investigation focuses on birth cohorts 1964 to 1972 only. This approach gives us the opportunity to observe the cohorts starting at age 20 until at least age 45. This restriction is important as we know in the German context that only lifetime earnings up to age 45 and older are sufficient proxies for complete lifetime earnings up to age 60 (Bönke et al., 2015). Second, we exclude individuals who were only observed prior to turning 30 since labor market patterns of individuals in their twenties are very unstable and could yield a life cycle bias (e.g., Brenner, 2010; Bönke et al., 2015; Haider and Solon, 2006). Further, the probability of observing the highest educational attainment accurately increases significantly with age 30 and older (Autorengruppe Bildungsberichterstattung, 2018) and observing the true educational attainment is crucial as education levels and earnings patterns over the work life are highly correlated (e.g., Bhuller et al., 2011; Bönke et al., 2015; Brunello et al., 2017). Third, we also exclude individuals without at least two consecutive observation years in the SOEP. Otherwise, no panel information is available and a distinction between individual short- and long-term labor shocks would not be possible. After eliminating those observations, we are left with a sample of 3,315 women and 3,212 men across birth cohorts 1964 to 1972 (see Appendix B, Table B.1) for the dynamic microsimulation by regression.

3.1.1 Dynamic Microsimulation Model

We apply a dynamic microsimulation by regression model to fill in missing information in non-observed years based on the individual's employment biography and socio-economic characteristics. To exploit our data to its fullest extent, we use both forward- and backward-looking simulations. The simulation starts either at an individual's first or last observed year in the data. As shown in Figure 3, we impute the missing variables in time $t+1$ or $t-1$ by running the regressions for our dynamic microsimulation in two consecutive steps: First, missing observations of marital status, fertility (i.e. number of children) and partners are simulated in the Family Module (Module 1). Second, the obtained information from Module 1 is used among others to simulate individuals' labor market information in the Labor Market Module (Module 2). Completing both modules yields the successful imputation of all relevant information in time $t + 1$ or $t - 1$. Afterwards, the process moves forward to the simulation of the next years, i.e. $t + 2$ or $t - 2$, $t + 3$ or $t - 3$, and so on.

The simulation ends after reaching 1984 in the backward looking and 2017 in the forward-looking process. We obtain a full dataset without any missing earnings or family information between 1984 and 2017.

Figure 3: Dynamic Microsimulation by Regression



In addition, investigating complete lifetime patterns for our sample requires us to extend our simulation for 15 additional years until 2032, when the youngest birth cohort 1972 turns 60. The prediction of employment biographies after 2017 are based on regression parameters of observed individuals from older cohorts, while we assume that general labor market characteristics (e.g., unemployment rate) remain stable after 2017. We also account for differences in trends using cohort and age fixed effects in our regressions. Nevertheless, this prediction comes naturally with a certain level of uncertainty due to the assumption that trends remain stable – an assumption that neglects pandemic related labor market effects. The simulation ends when all missing information between 1984 and 2032 is simulated.

Within each module, the simulation of variables is based on estimating transition probabilities between two years, e.g., if marital status changes from year t to $t+1$. The estimation of a change of a variable j between two periods is then implemented by using a random process (e.g., Neufeld, 2000; Plümper and Troeger, 2007; Zucchelli et al., 2010): For each individual observation i we simulate the transition probability from time t to $t+1$ or $t-1$ and then draw a random number N_{it} from a uniform $[0, 1]$ distribution. If the calculated transition probability P_{it} is larger than the drawn random number N_{it} ($P_{it} > N_{it}$), a transition occurs. In contrast, no transition takes place if $P_{it} \leq N_{it}$. Therefore, high transition likelihoods do not always induce actual transitions and even low transition probabilities may still lead to transitions. This approach helps to account for the uncertainty that comes with a simulation. Additionally, we use a Monte Carlo simulation approach to test the robustness of our results (see Appendix B, Figures B.3 and B.4). The results of the Monte Carlo simulation confirm the high robustness of our simulation outcomes.

Next, we will give brief summaries about both simulation modules. Detailed information on all the regression models of every simulation step can be found in Table B.2 in Appendix B.

3.1.2 Module 1: Family Module

Empirical evidence shows that family background strongly influences women's labor market behavior (Kleven and Landais, 2017). Therefore, we need information on individual's family background before simulating earnings for non-observed years. All individuals in our sample completed entry questionnaires including questions on marital status and, if applicable, birth years of children before entering the survey; this allows us to reconstruct full family histories. Consequently, missing data occurs exclusively after individuals left the survey. This eliminates the necessity of the backward looking simulation component in this module. In addition, we also observe most women at older ages so only around 20 percent of child information must be simulated.

The Family Module then consists of two steps: predicting marital status, including a partnering module when necessary, and predicting births of children for individuals with missing information. First, we run logistic regressions separately by gender s (Female or Male) and marital status m (Single or Partnered) in year t to predict the individual transition probability $p^{married}$ to change the marital status from year t to the missing year $t+1$:

$$p^{married}_{m,s,t+1} = \beta_0 + \beta_1 X_{m,s,t} + \epsilon_{m,s,t}, \quad E(\epsilon_{m,s,t}) = 0, m \in \{S, P\}, s \in \{F, M\}, t \in [1984, 2017] \tag{3}$$

The regression consists of a set of explanatory variables X_t including socio-economic characteristics (e.g., education, age, migration background) and labor market behavior (e.g., employment status). In addition, we control for the number of years that an individual's marital status has remained unchanged until year t. Table B.2 (see Appendix B) gives a detailed overview about all covariates included in each regression-based simulation step.

Recall that if $P_{it} \leq N_{it}$, the marital status stays the same and if $P_{it} > N_{it}$, the marital status changes. Therefore, this simulation step has four possible outcomes: First, a person who is single in year t can remain single in $t+1$. Second, married individuals can stay married. Here we assume that their partners remain the same. Third, married individuals in period t can get divorced and become single in $t+1$.[13] And fourth, singles in year t can get married in t+1. In this last case, we run a Partner Module to assign a partner.[14] This allows us to account for partners' characteristics when simulating family and labor market decisions. Using Mahalanobis distance matching we identify five "best" partners based on age, education and region for each observation. We then randomly assign one of the five potential partners to

[13]In this case we assume that the children stay with the mother. Empirical evidence by the Federal Statistical Office (2018) supports this assumption: The share of single fathers is only around 10 to 13 percent since 1997.

[14]For a few married individuals in our data, we cannot observe partner information since the partner did not participate in the survey, e.g., because she or he refused. In those cases, we also run the Partner Module as a preparation step before starting the Family Module.

the individual. Our matching procedure is not unique, i.e., one individual can serve multiple times as a "donor" for partner characteristics. In this way, we ensure a sufficient pool of potential partners.[15]

Next, we simulate whether a woman will give birth to a child in the next non-observed period $t+1$ by marital status m:

$$p^{birth}_{m,t+1} = \beta_0 + \beta_1 X_{m,t} + \epsilon_{m,t}, \quad E(\epsilon_{m,t}) = 0, m \in \{S, P\}, t \in [1984, 2017]. \tag{4}$$

Again, X_t represents a set of explanatory variables including socio-economic characteristics like information on existing children and labor market information. The simulation is similar to the approach described in the simulation of the marital status. Afterwards, the information on an individual's number of children is then updated accordingly. In contrast to our marriage simulation, births are only simulated for women. Children are then attached to men depending on women's family background.

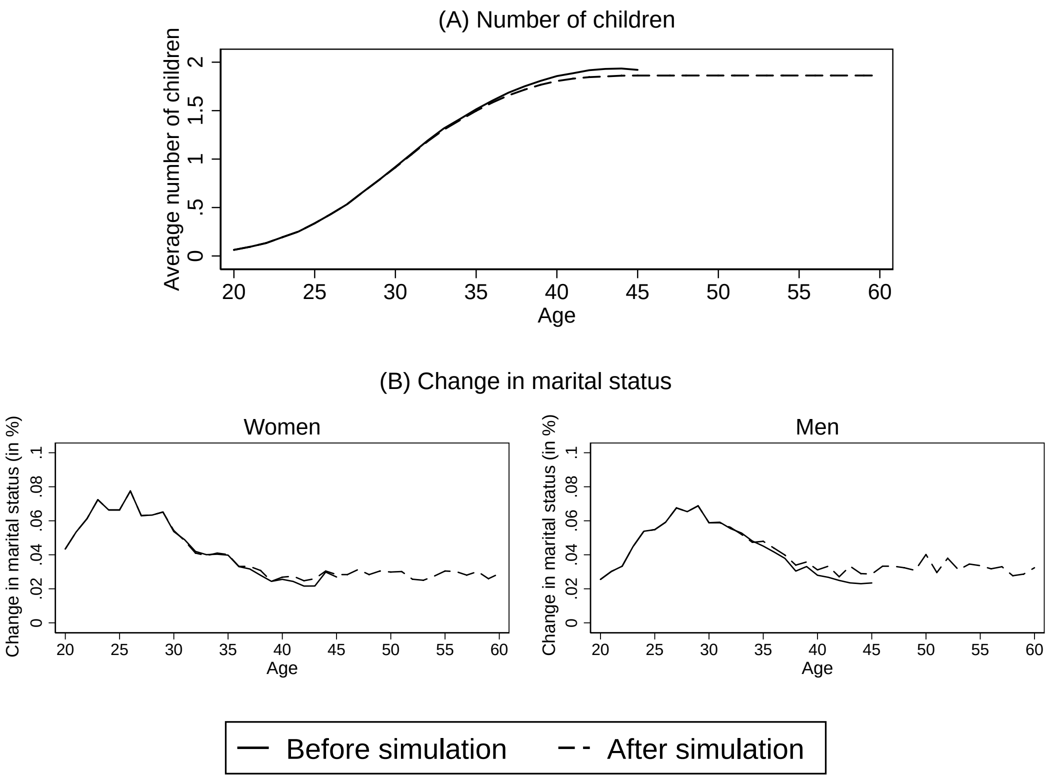
Since we estimate transition likelihoods for $t+1$ by using information available in period t, the likelihood of a change of the marital status or a childbirth in $t+1$ do not influence the transition probability of one another. Therefore, the order in which we implement fertility and marital transitions is irrelevant and does not alter our results.

Completing the Family Module for years 1984 to 2032 results in a sample with full information on family characteristics. Figure 4 shows that our simulated data (dashed line) replicates the initial distributions before the simulation (solid line) very accurately. In Panel A, the women's average number of children increases strongly

[15]It is important to note that although we can match individuals with their partners for observed years in our data, we do not generate a household perspective. Therefore, this approach is not problematic as we perform all our estimations separately for partners and only include the partner's characteristic as explanatory variables.

until age 35. Then, the growth rate slows down and comes to a natural stop between ages 40 and 45 due to biological reasons. Panel B displays the percentage change in marital status by age. Obviously, trends for both men and women follow the same trend over life. Most changes in marital status happen in the beginning of life. Between ages 35 and 45, only around 3 percent of men and women get married or divorced.

Figure 4: Family Information Before and After Simulation



Notes: Panel A shows the average number of children of women by age before and after the simulation. Panel B demonstrates the share of individuals in our sample changing their marital status before and after the simulation.

Source: Own calculations based on SOEP v.35 (1984-2018).

3.1.3 Module 2: Labor Market Module

The Labor Market Module generates complete information on an individual's employment biography through five stages: labor market participation, employment status, type of work arrangement (full-time or part-time), annual working hours and annual earnings. In this module, we use both forward and backward simulation as the introductory survey questionnaires do not allow us to construct sufficient work histories. Our model description will focus on the forward-looking simulation component. However, the backward-looking part of the simulation follows the same methodology.

In general, the logic and structure of this module is very similar to our approach in the Family Module. We start with the estimation of $p^{lmp}_{(m,t+1)}$, the probability for an individual of marital status m to change the labor market participation lmp from year t to year $t+1$. The labor market participation dummy variable is equal to 1 if individuals are unemployed or employed and equal to 0 if they are not attached to the labor market (e.g., due to parental or sick leave). We run the estimation separately by gender s and marital status m:

$$p^{lmp}_{s,m,t+1} = \beta_0 + \beta_1 p^{lmp}_{s,m,t} + \beta_2 p^{lmp}_{s,m,t-1} + \beta_3 X_{s,m,t} + \epsilon_{s,m,t},$$

$$E(\epsilon_{s,m,t}) = 0,\ s \in \{F, M\},\ m \in \{S, P\},\ t \in [1984, 2017]. \tag{5}$$

$X_{(s,m,t)}$ is again a vector of control variables with socio-economic characteristics like marital status, partner's earnings and their own labor market information. Further, we include lagged dependent variables to account for path dependencies over the work life while still modelling a dynamic data generating process.[16] If individuals are recorded as not participating in year $t+1$, we directly record their earnings as

[16] We only include two lags as more lags would decrease our sample size. For this estimation strategy, we are able to include all individuals that have at least two observation years in the SOEP. Each additional lag would restrict our sample to individuals with more observed years.

zero for $t+1$ and do not include them in the subsequent steps. For individuals who are active in the labor market, we next run a regression to estimate the probability to change their employment status $p^{emp}_{(s,m,e,t+1)}$ (employed/unemployed) from year t to year $t+1$. The following model is run separately by gender s, marital status m and employment status e:

$$p^{emp}_{s,m,e,t+1} = \beta_0 + \beta_1 p^{emp}_{s,m,e,t} + \beta_2 p^{emp}_{s,m,e,t-1} + \beta_3 X_{s,m,e,t} + \epsilon_{s,m,e,t},$$

$$E(\epsilon_{s,m,e,t}) = 0,\ s \in \{F, M\},\ m \in \{S, P\},\ e \in \{0, 1\},\ t \in [1984, 2017]. \qquad (6)$$

Once more, the regression contains of a set of explanatory variables $X_{(s,m,e,t)}$ including information of the family and the socio-economic background. Also included in the control vector is the work history of individuals. To this end, we measure work experience by years of full-time work, part-time work and years without any work until year t to account for the different levels of labor market experience.

Individuals recorded as unemployed in year $t+1$ after this first regression step receive zero earnings in $t+1$ and are excluded from further estimations. For all employed individuals, the dynamic microsimulation moves forward with a logistic regression simulating if individuals worked full- or part-time in year $t+1$. Following, we estimate the probability of changing full-time or part-time arrangements from year t to year $t+1$:

$$p^{wt}_{(s,m,t+1)} = \beta_0 + \beta_1 p^{wt}_{s,m,t} + \beta_2 p^{wt}_{s,m,t-1} + \beta_3 X_{s,m,t} + \epsilon_{s,m,t},$$

$$E(\epsilon_{s,m,t}) = 0,\ s \in \{F, M\},\ m \in \{S, P\},\ t \in [1984, 2017]. \qquad (7)$$

Again, $X_{(s,m,t)}$ includes the usual control variables in addition to the labor market history. We can now move on to estimate the precise number of annual working hours in $t+1$ separately for part-time and full-time workers. We use an OLS regression model following the same logic as the earnings regression model as introduced in Equation 8.[17]

Finally, we use an earnings regression to estimate the annual earnings $y_{(s,m,t+1)}$ by gender s and marital status m:

$$y_{(s,m,t+1)} = \beta_0 + \beta_1 y_{s,m,t} + \beta_2 y_{s,m,t-1} + \beta_3 X_{s,m,t} + \epsilon_{s,m,t},$$

$$E(\epsilon_{s,m,t}) = 0,\ s \in \{F, M\},\ m \in \{S, P\},\ t \in [1984, 2017]. \tag{8}$$

$X_{(s,m,t)}$ now includes information about the work history in years of full-time work, part-time work or unemployment, working hours in t and, if applicable, partner and child information. All earnings are in prices of 2015. After completing all five steps of the Labor Market Module between 1984 and 2017, all individuals have complete employment and earnings information for previously unobserved years. Afterwards, we continue the simulation until 2032 to obtain complete biographical data up to age 60.

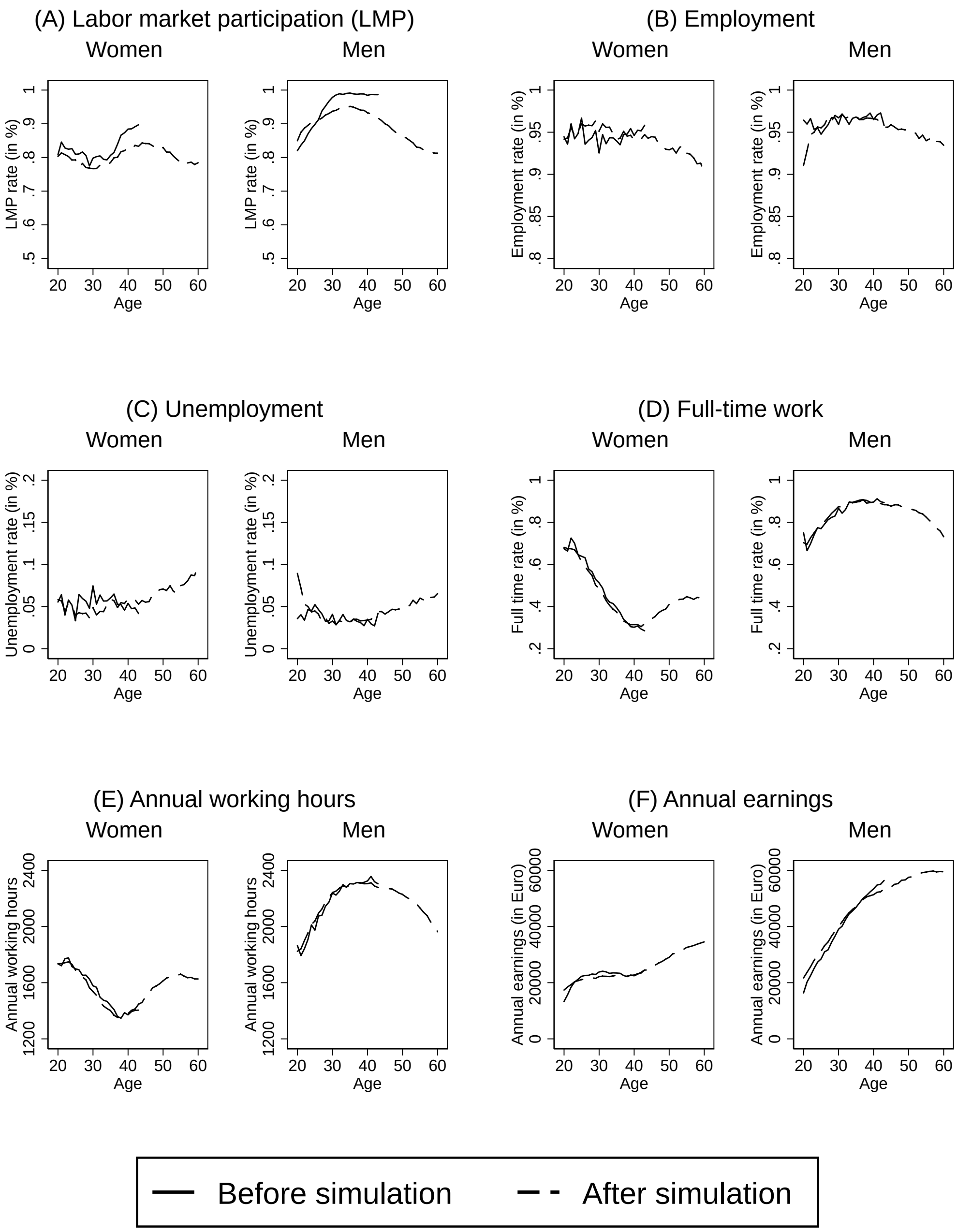
Figure 5 shows that our simulated data (dashed line) replicates the original SOEP data (solid line) well, particularly for Panel D (Full-time work), Panel E (Working hours) and Panel F (Earnings). Panel A (Labor Market Participation), Panel B (Employment) and Panel C (Unemployment) show small deviations. Most of these differences occur in the beginning of the work life. These differences do not necessarily diminish the quality of our microsimulation for the following two reasons: First, our sample restriction to individuals observed at least once at age 30 or older leads to less observations in individuals' early twenties. As a result, our SOEP sample before the simulation is not very reliable for this age range due to a small sample size, and therefore comparisons may be misleading. Second, earnings in the early work life only account for a very small portion of lifetime earnings so

[17] Again, see Table B.2 in Appendix B for more detailed information.

they do not play a major role in shaping lifetime outcomes. After the completion of both modules of our dynamic microsimulation by regression, we obtain all relevant labor market and household information for birth cohorts 1964 to 1972 from age 20 to 60 to proceed with our lifetime analysis.[18] Overall, the simulated data mirrors the data patterns before simulation and our simulation results are robust. Additional robustness checks based on a Monte Carlo simulation approach and the simulation of pseudo-missings can be found in Appendix B.1.

[18]Our sample after the microsimulation is significantly different from our original SOEP sample. Therefore, we cannot use the longitudinal weights initially provided by the SOEP. To maintain representativeness, we therefore use census data (Mikrozensus) to reweight our sample with regard to cohort, age, family and labor market information. The Mikrozensus is considered highly representative for Germany, covering about one percent of the entire German population through mandatory participation.

Figure 5: Labor Market Information Before and After Simulation



Notes: Only employed individuals are considered. Panels D - F do not include values of zero annual earnings. Cohorts 1940-1979, weighted sample.

Source: Own calculations based on SOEP v.35 (1984-2018).

3.2 Lifetime Analysis

Although we have already shown that women face lower hourly wages and annual earnings, and are less active on the labor market, the cross-sectional analysis only reveal a snapshot of an individual's employment biography. A cross-sectional analysis does not reveal how advantages or disadvantages might add up or balance out over the lifecycle. For a better understanding of when and how in life the gender gap develops, we investigate differences in accumulated earnings over the lifecycle for birth cohorts 1964 to 1972 using their complete biography data from age 20 to 60 obtained from our microsimulation. To analyze the accumulation of earnings over the work life we follow Bönke et al. (2015) and use the "up-to-age-X" (UAX) concept. UAX earnings refer to accumulated earnings up to a certain age X. We define lifetime earnings as UA60 earnings.

3.2.1 Gender Gap in Lifetime Earnings

To analyze the gender gap in lifetime earnings, we now focus on nonlogarithmic incomes rather than logarithmic incomes as used in the Oaxaca Blinder decomposition[19] in Section 2. Using logarithmic incomes would lead to the exclusion of zero earnings and, thus, periods of inactivity.[20] Since especially women accumulate periods of inactivity over life through motherhood and childbearing, times of inactivity without any earnings play a crucial role in assessing the lifetime earnings gender gap and need to be included in this analysis.

The gender gap G in the labor market outcome variable L (here: hourly wages,

[19] As stated in Section 2, the Oaxaca Blinder decomposition is based on an OLS regression model using log hourly wage and log annual earnings.

[20] The inverse hyperbolic sign (ihs) transformation represents an alternative concept. In contrast to the logarithmic transformation, it is also defined for negative and zero values (e.g., Burbidge et al., 1988; Pence, 2006) (e.g., Burbridge et al. 1988, Pence 2006). Due to these advantages, it is primarily used in the literature on wealth distributions (e.g., Pence, 2006; Grabka et al., 2015; Sierminska et al., 2019). However, the literature on gender differences does not use this concept and to maintain comparability of our results, we stay in line with this literature strand.

annual earnings, UAX earnings) in percent for men m and women f at age x is now defined as:

$$G_x = [(\overline{L}_{m,x} - \overline{L}_{f,x})/\overline{L}_{m,x}] \times 100 \ . \tag{9}$$

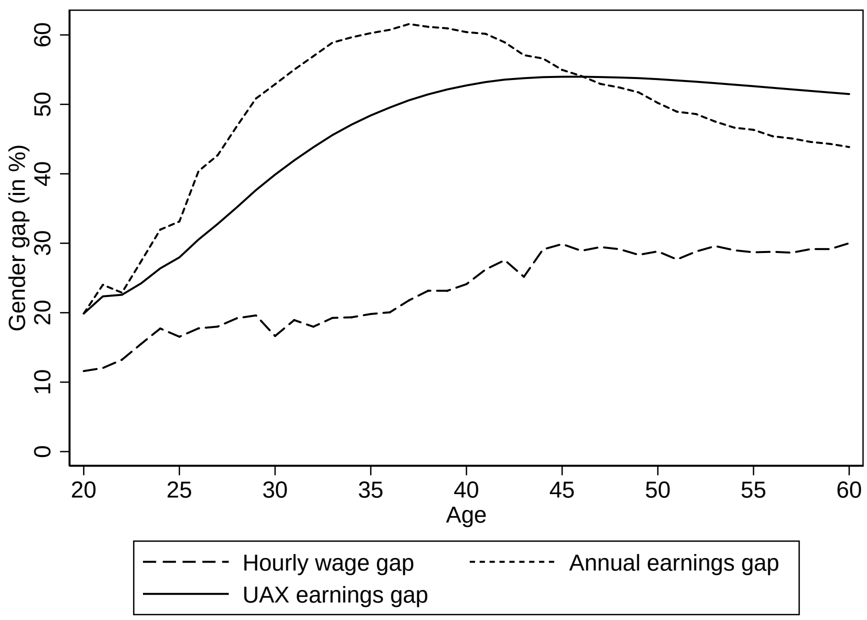
Based on our new sample obtained from the microsimulation, Figure 6 shows the gender gaps in hourly wages, annual earnings and UAX earnings for ages 20 to 60 for birth cohorts 1964 through 1972. As expected, despite the same trend we see several differences when we compare the gender gaps in hourly wages and annual earnings using this microsimulation sample to our results based on the cross-sectional sample discussed in Section 2.

At early ages, the difference in pay by gender is rather low but then increases steadily until retirement. However, we can observe differences in levels which are driven by the more confined cohort restriction in our microsimulation sample and the varying definition of the gender gap (logarithmic vs. non-logarithmic income). Comparing the gender gaps in annual earnings reveals more pronounced differences between the cross-sectional and lifetime approach. First, the inversely U-shaped gender gap in annual earnings in Figure 6 is significantly larger than the gender gap shown in Figure 2. This difference is largely driven by the inclusion of inactive labor periods with zero earnings in this lifetime analysis, while we excluded those in our cross-sectional analysis in Section 2.[21] Including periods with zero earnings leads to a decline in women's average earnings, and thus to an increase in the gender gap. Naturally, this difference is especially pronounced in the years of family formation since women are more often inactive due to child rearing. Second, in contrast to the gender gap estimated using the cross-sectional sample, Figure 6 shows a pronounced decline of the gender gap in annual earnings between ages 40

[21]See Figure C.1 in Appendix C for a direct comparison of the gender gap in annual earnings when including or excluding individuals with zero earnings.

and 60. Again, this difference is driven by the different composition of our two samples. While the cross-sectional sample includes all birth cohorts 1940 to 1979, the lifetime sample is restricted to younger cohorts. Due to the higher labor market participation rates for women of younger cohorts, the gender gap in annual earnings declines again before retirement once we restrict our sample to younger cohorts, because our simulation assumes that women of younger cohorts would reenter the labor market after inactive labor market periods more often.

Figure 6: Gender Gaps in Wages, Annual Earnings and UAX Earnings over Life



Notes: Individuals with zero UAX earnings are included in the calculation. For annual earnings, employed and unemployed individuals are considered. For hourly wages, only employed individuals are considered. Cohorts 1964-1972.

Source: Own calculations based on SOEP v.35 (1984-2018).

Finally, the solid line in Figure 6 shows the gender gap in UAX earnings as the sum of the annual earnings up to age X. Ultimately, the UA60 earnings coincide with our definition of lifetime earnings. Hence, the higher the age X, the closer UAX earnings are to lifetime earnings. At the beginning of the work life, women earn on average 20 percent less than men do. The difference in earnings accumulates over the life course and increases to a gender gap in UA40 earnings of 52.7 percent. After that, the gap remains stable, which results in a gender gap in lifetime earnings of 51.5 percent (UA60). At this point in life, women have earned on average around 732.000 Euro in 2015 prices—slightly less than half of the average income that men were able to accumulate (1.510.000 Euro).[22]

The evolution of the gender gap in UAX earnings is by construction driven by the gender gap in the annual earnings curve. UAX earnings are less volatile since the marginal effect of adding an additional year of annual earnings to the UAX earnings decreases with increasing age. Hence, the gender gaps in annual and UAX earnings both experience large growth until age 40, but when the gender gap in annual earnings declines again, the UAX gender gap remains at its high level.

The profound difference in lifetime earnings is largely the result of differences in the extensive and intensive margin of labor supply of women over their life. One can discuss how labor supply is influenced by own decisions or forced by personal and social circumstances. Previous studies have shown a strong relationship between gender gaps in income and children (e.g., Angelov et al., 2016; Kleven and Landais, 2017; Adda et al., 2017). This can be partially explained by the close connection between women's labor market decisions and the number of children they have (Kühhirt and Ludwig, 2012; Ejrnæs and Kunze, 2013). In line with these studies, we also find that mothers face higher earning losses with every additional child, while fatherhood does not seem to affect men's earnings. Hence, observed earnings differences between childless women and men are smallest and grow wider with every additional child (see Figure C.2, Appendix C). This observation also holds true when we analyze the evolution of UAX earnings by number of children (Figure

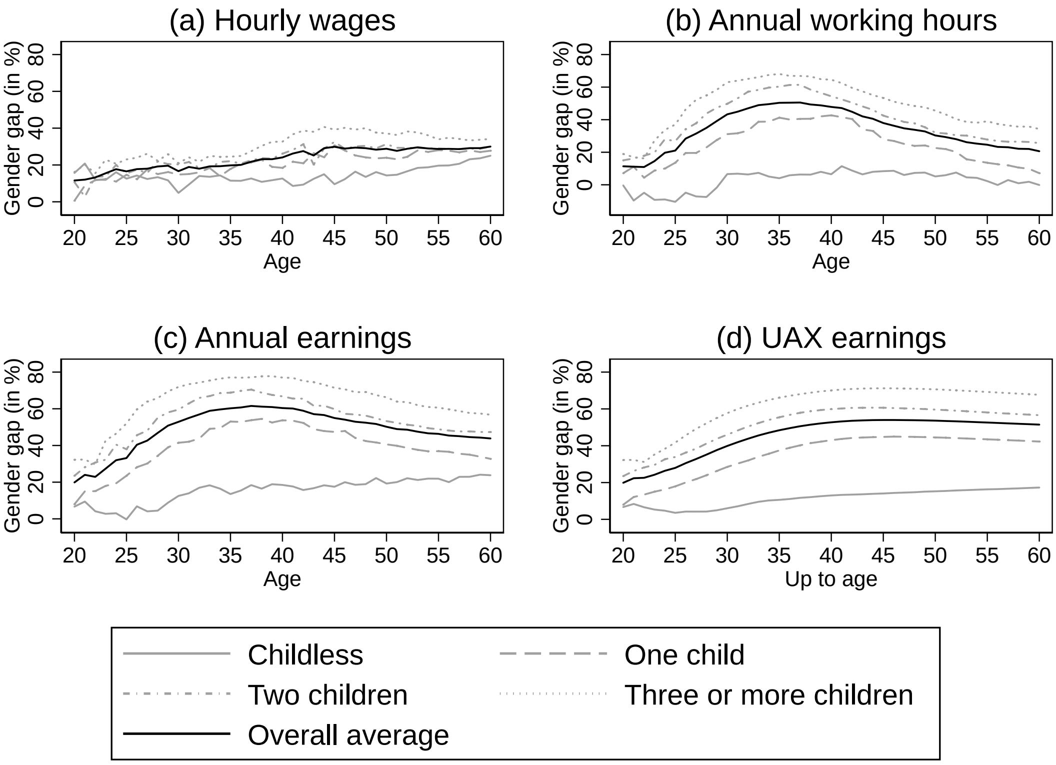
[22]Compare Figure C.2 and Figure C.3 for the distribution of annual earnings and UAX earnings by men and women over the work life.

C.3, Appendix C).

Figure 7 shows the gender gap in hourly wages (a), the gender gap in hours worked (b), the gender gap in annual earnings (c) and the gender gap in UAX earnings (d) over the life cycle by number of children. In the beginning, the gender gap in hourly wages shows only small differences for women with and without children but widens over the life cycle. In Section 2, we have shown that this is mainly explained by the lesser work experience women with children gain over their life courses. The gender gap in annual earnings clearly differs by the number of children throughout the entire life cycle (see Figure 7c), exacerbating the gap in hourly wages mainly due to mother's lower intensive margin of work (see Figure 7b).

The gender gap in lifetime earnings also increases with the numbers of children. While childless men and women experience a gender gap of 17.3 percent, the gap is significantly higher for men and women with three or more children (68.0 percent at age 60). The significant widening of the gender gap between up to age 35 earnings thereby coincides with the increase in the cross-sectional gender gaps in annual hours worked, and consequently annual earnings. These results are in line with existing studies finding evidence for motherhood penalties and fatherhood premiums (e.g., Budig and England, 2001; Killewald and Gough, 2013; Killewald and García-Manglano, 2016). Therefore, descriptive evidence clearly hints that motherhood might be a key driver of gender earnings inequality over the life cycle.

Figure 7: Gender Gaps over the Lifecycle by Children



Notes: Number of children refers to the total number at age 60. Individuals with zero annual and UAX earnings are included in the calculation.

Source: Own calculations based on SOEP v.35 (1984-2018).

3.2.2 Counterfactual Analysis

In the last step, we want to determine which part of the observed gender gap in lifetime earnings can be associated with differences in the distribution of characteristics across gender and which part is associated with differences in returns to characteristics. To investigate this issue further, we will predict counterfactual lifetime earnings for women in the following two steps.[23]

First, we take the earnings regression results from our microsimulation model,

[23]Due to restrictions of our simulated sample, we cannot estimate an Oaxaca Blinder decomposition to decompose the gender gap in UAX earnings.

estimated for male m and female f individuals separately:

$$\hat{y}_{s,t+1} = \hat{\beta}_{0,s} + \hat{\beta}_{1,s} y_{s,t} + \hat{\beta}_{2,s} y_{s,t-1} + \hat{\beta}_{3,s} X_{s,t}, \quad s \in \{F, M\} \text{ and } t \in [1984, 2017] \quad (10)$$

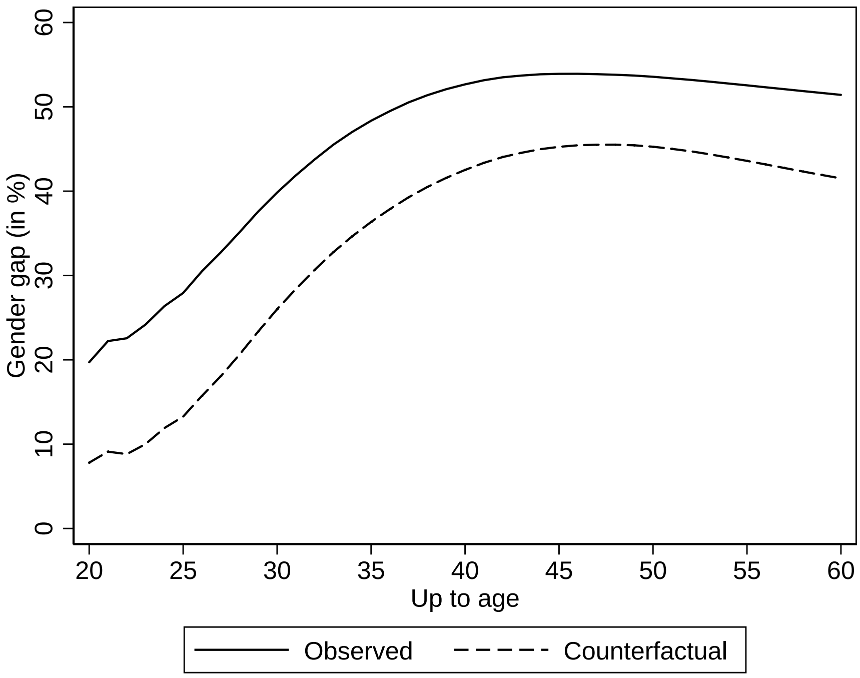
Second, we then estimate women's counterfactual annual earnings $\hat{y}_f^C$ by using the coefficients obtained from the male regression model in the women's Mincer earnings regression:

$$\hat{y}_{f,t+1}^C = \hat{\beta}_{0,m} + \hat{\beta}_{1,m} y_{f,t} + \hat{\beta}_{2,m} y_{f,t-1} + \hat{\beta}_{3,m} X_{f,t}, \quad t \in [1984, 2017] \quad (11)$$

Women's counterfactual annual earnings in year t then represent the salary women would have earned if their characteristics were as equally rewarded as men's. Adding up the counterfactual annual earnings for each woman over the life course then yields women's counterfactual UAX earnings. As a result, all differences displayed in the counterfactual gender lifetime earnings gap are solely based on different characteristics for men and women and not by different returns to characteristics.

Figure 8 compares the observed and counterfactual gender gaps in UAX earnings. The difference between the truly observed and the counterfactual gender gap can be interpreted as the unexplained part of the gender gap in UAX earnings (adjusted gender gap). In the beginning of the work life, the difference between both gaps shown in Figure 8 is 12.1 percentage points. Therefore, in early years, approximately half of the gender gap in UAX earnings is due to a different allocation of characteristics and half is due to a different reward or payment of characteristics. The adjusted gender gap then increases to about 14.8 percent for UA30 earnings and declines afterwards to 10 percent for lifetime earnings (UA60). Thus, until the years of family formation, the unexplained difference between women's and men's pay grows, whereas it declines towards retirement. Overall, 80 percent of the observed gender lifetime earnings gap of 51.5 percent at age 60 can be explained by a different

Figure 8: Counterfactual Estimation of the Gender Lifetime Earnings Gap



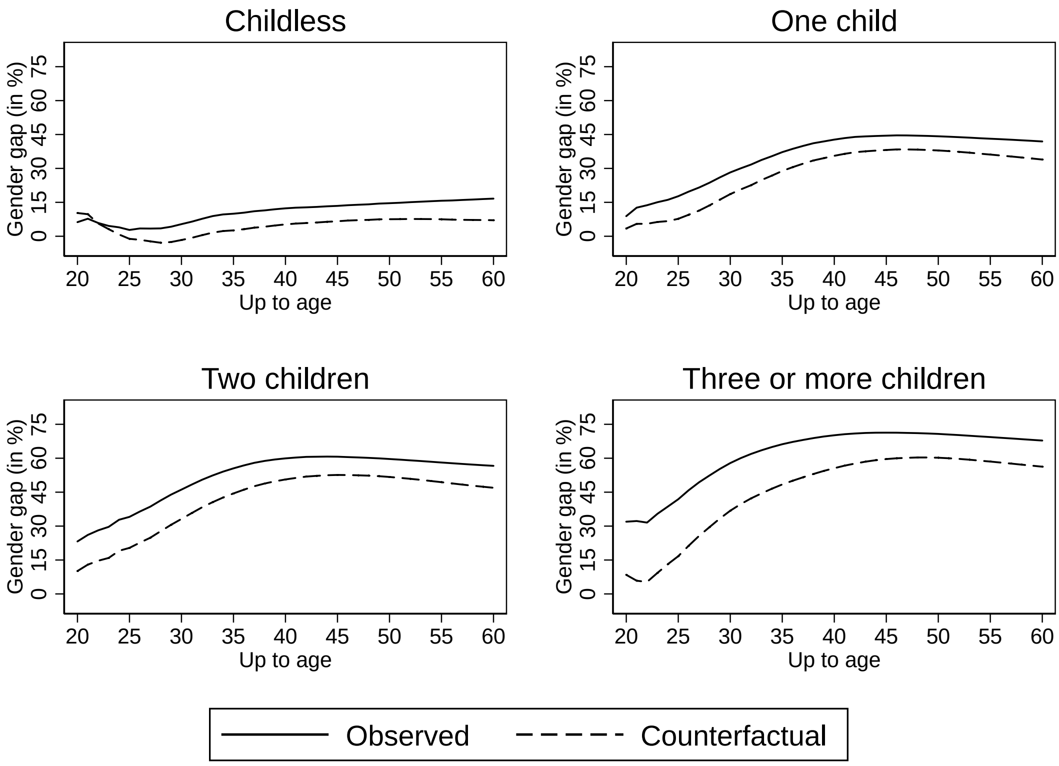
Notes: Estimated and counterfactual gender gap in UAX earnings. Gender gaps in accumulated earnings are earnings up to a given age. Individuals with zero UAX earnings are included in the calculation.

Source: Own calculations based on SOEP v.35 (1984-2018).

distribution of labor market characteristics of men and women. Consequently, the remaining one fifth of the observed gender lifetime earnings gap of 51.5 percent at age 60 is due to a less favorable reward for women's labor market characteristics, leading to an overall adjusted gender lifetime earnings gap of 10 percent. The evolution of the adjusted gender gap indicates that rewards are least favorable for women in the first half of their work life. As this is the main time for family formation, this might be due to either a sorting of women into worse positions to gain more flexibility or the labor market rewarding women less favorably during this time due to the higher risk of inactivity periods.

Next, we want to investigate how motherhood influences the adjusted gender gap in lifetime earnings other than through their labor market characteristics alone..

Figure 9: Counterfactual Estimation of the Gender Lifetime Earnings Gap by Number of Children



Notes: Estimated and counterfactual gender gaps in UAX earnings. Gender gaps in accumulated earnings are earnings up to a given age. Individuals with zero UAX earnings are included in the calculation.

Source: Own calculations based on SOEP v.35 (1984-2018).

Hence, Figure 9 compares the observed and counterfactual gender gaps by the number of children. As already shown in Figure 7d, the observed gender gap in lifetime earnings is lowest for childless women and increases strongly with the number of children women have. But how much of the observed gender gap in lifetime earnings of women with and without children can be explained by a different distribution of characteristics, and what is the influence of the role of motherhood on the adjusted gender gap in lifetime earnings?

Using German data, this paper shows for the first time that in stark contrast to the observed gender gap in UAX earnings, the adjusted gender gap only slightly differs by the number of children women and men have. The difference between women with and without children amount to only three percentage points, with mothers of three or more children facing the highest adjusted gender gap in lifetime earnings with 11.4 percent. Hence, the large differences in the observed gender gaps of women with and without children are mainly driven by the different accumulation of characteristics rather than an additional unexplained penalty of motherhood.

Although the adjusted gender gap in lifetime earnings does not differ much for women with and without children, interestingly, the evolution over the work life until age 60 does differ. For women with children, the adjusted gender gap in UAX earnings remains mostly stable over the work life, showing the same growth patterns for both the observed and the counterfactual gender gaps. Consequently, the increase in the observed gender gap in UAX earnings over the work life for women with children is driven by the accumulation of unfavorable characteristics (e.g., less work experience). In contrast, the analysis reveals a very different development for women without children. At the beginning of the work life, the adjusted gender gap in UAX earnings is nonexistent or even negative for women without children. This indicates that their characteristics are rewarded the same as or even slightly better than the characteristics of their male counterparts. But with increasing age, the adjusted gender gap in UAX earnings grows for women without children, peaking at the end of their work life. One possible cause for this different development could be that although women without children start off in similar positions as men, over their careers they are promoted less frequently into higher positions; a phenomenon often referred to as the glass ceiling. Another possible cause could be that these women are selecting themselves into sectors or occupations with lower average earnings growth potential (e.g., care sector).

Overall, we show that the difference in the gender gap in lifetime earnings by motherhood is largely driven by different characteristics women accumulate over their work life. Our results in Section 2 and Figure 7b indicated that these differences are primarily due to fewer working hours and less work experience women with children accumulate over their work life. Nevertheless, at the end of the work life all women face an adjusted gender gap in lifetime earnings of around 10 percent; this is only due to the less favorable reward for their characteristics on the labor market compared to their male counterparts which cannot be explained by our models.

4 Conclusion

This paper underlines the importance of accounting for the biographical dimension when analyzing gender inequalities. Our results show that cross-sectional gender differences are persistent over the work life. Comparing multiple dimensions of cross-sectional gender differences, we find that the gender gap in hourly wages is less than half the amount of the gender gap in annual earnings; peaking at age 40, the gender gap in hourly wages is 0.343 log points compared to 0.829 log points for annual earnings. Using an Oaxaca Blinder decomposition, we show that especially the gender gap in annual earnings can largely be explained by the extensive and intensive margin of labor with women having less work experience and working less hours. We then applied a dynamic microsimulation model to obtain full lifetime earnings data including family background information. Using our simulated data, we observe a gender gap in lifetime earnings of 51.5 percent, increasing with the number of children women have. While childless women face an average gender gap in lifetime earnings of 17.2 percent, mothers with three or more children experience a gap of 67.7 percent. Next, we used the coefficients from the male earnings regression simulation model to estimate women's counterfactual earnings. As a result, all differences remaining were solely based on different characteristics of men and women and not by different returns to characteristics. The difference between the truly observed gender gap and the counterfactual gap then yielded the adjusted gender gap in lifetime earnings of 10 percent. This means that women earn on average 10 percent less than men over their lifetime due to a different reward for their characteristics in comparison to men. We find that in stark contrast to the observed gender gap in UAX earnings, the adjusted gender gap only differs slightly by the numbers of children women and men have.

The documented gender inequalities in lifetime earnings are high and concerning for a variety of social and economic reasons. Less financial opportunities for women, and especially mothers, might create unhealthy dependency structures within households. High opportunity costs for having a family may lead to less women wanting children, or vice versa less women aiming for higher career goals despite their large investments in education. Lower lifetime earnings also result in significantly lower pensions and consequently a higher risk of poverty among elderly women. Furthermore, it is still not clear to what extent women's lower labor market participation is voluntarily or a result of unfavorable work and social conditions. This underlines the importance of further research on the dynamics of gender differences over the life course.

References

Adda, J., C. Dustmann, and K. Stevens (2017). The career costs of children. *Journal of Political Economy 125*(2), 293–337. Publisher: University of Chicago Press Chicago, IL.

Altonji, J. G. and R. M. Blank (1999). Race and gender in the labor market. In *Handbook of Labor Economics*, Volume 3, pp. 3143–3259. Elsevier.

Anderson, D. J., M. Binder, and K. Krause (2002). The motherhood wage penalty: Which mothers pay it and why? *American economic review 92*(2), 354–358.

Angelov, N., P. Johansson, and E. Lindahl (2016). Parenthood and the Gender Gap in Pay. *Journal of Labor Economics 34*(3), 545–579. Publisher: University of Chicago Press Chicago, IL.

Autorengruppe Bildungsberichterstattung (2018). *Bildung in Deutschland 2018 Ein indikatorengestützter Bericht mit einer Analyse zu Wirkungen und Erträgen von Bildung.* Bielefeld: wbv Media.

Beblo, M. and E. Wolf (2002). Die Folgekosten von Erwerbsunterbrechungen. *Vierteljahrshefte zur Wirtschaftsforschung 71*(1), 83–94. Publisher: Berlin: Duncker & Humblot.

Bertrand, M., C. Goldin, and L. F. Katz (2010). Dynamics of the gender gap for young professionals in the financial and corporate sectors. *American economic journal: applied economics 2*(3), 228–55.

Bhuller, M., M. Mogstad, and K. G. Salvanes (2011). Life-Cycle bias and the returns to schooling in current and lifetime earnings. SSRN Scholarly Paper ID 1774762, Social Science Research Network, Rochester, NY.

Björklund, A. (1993). A comparison between actual distributions of annual and lifetime income: Sweden 1951–89. *Review of Income and Wealth 39*(4), 377–386. Publisher: Wiley Online Library.

Blau, F. D. and L. M. Kahn (2017). The gender wage gap: Extent, trends, and explanations. *Journal of Economic Literature 55*(3), 789–865.

Blinder, A. (1973). Wage discrimination: Reduced form and structural estimates. *Journal of Human Resources 8*, 436–55.

Bönke, T., G. Corneo, and H. Lüthen (2015). Lifetime Earnings Inequality in Germany. *Journal of Labor Economics 33*(1), 171–208. Publisher: The University of Chicago Press.

Boll, C., M. Jahn, and A. Lagemann (2017). The gender lifetime earnings gap—exploring gendered pay from the life course perspective. *Journal of Income Distribution®*, 1–53.

Bonin, H., K. Reuss, and H. Stichnoth (2015). Life-cycle incidence of family policy measures in Germany: Evidence from a dynamic microsimulation model. Publisher: SOEPpaper.

Brenner, J. (2010). Life-cycle Variations in the Association between Current and Lifetime Earnings: Evidence for German Natives and Guest Workers. *Labour Economics 17*(2), 392–406.

Brown, J. R., J. L. Coronado, and D. Fullerton (2009). Is Social Security Part of the Social Safety Net? *Tax policy and the economy 23*(1), 37–72. Publisher: The University of Chicago Press.

Brunello, G., G. Weber, and C. T. Weiss (2017). Books are forever: Early life conditions, education and lifetime earnings in europe. *The Economic Journal 127*(600), 271–296. Publisher: Oxford Academic.

Budig, M. J. and P. England (2001). The Wage Penalty for Motherhood. *American Sociological Review 66*(2), 204–225. Publisher: [American Sociological Association, Sage Publications, Inc.].

Burbidge, J. B., L. Magee, and A. L. Robb (1988). Alternative transformations to handle extreme values of the dependent variable. *Journal of the American Statistical Association 83*(401), 123–127.

Coronado, J. L., D. Fullerton, and T. Glass (2011). The Progressivity of Social Security. *The B.E. Journal of Economic Analysis & Policy 11*(1). Publisher: De Gruyter Section: The B.E. Journal of Economic Analysis & Policy.

Correll, S., S. Benard, and I. Paik (2007). Getting a Job: Is There a Motherhood Penalty? *American Journal of Sociology 112*(5), 1297–1339. Publisher: The University of Chicago Press.

Ejrnæs, M. and A. Kunze (2013). Work and Wage Dynamics around Childbirth. *The Scandinavian Journal of Economics 115*(3), 856–877.

Federal Statistical Office (2017). Verdienste im Überblick. Technical report, Federal Statistical Office, Wiesbaden.

Federal Statistical Office (2018). Alleinerziehende in deutschland. Technical report, Federal Statistical Office, Wiesbaden.

Gallen, Y., R. V. Lesner, and R. Vejlin (2019). The Labor Market Gender Gap in Denmark: Sorting Out the Past 30 Years. *Labour Economics 56*, 58–67. Publisher: Elsevier.

Gangl, M. and A. Ziefle (2009). Motherhood, labor force behavior, and women's careers: An empirical assessment of the wage penalty for motherhood in Britain, Germany, and the United States. *Demography 46*(2), 341–369. Publisher: Springer.

Geyer, J. and V. Steiner (2014). Future public pensions and changing employment patterns across birth cohorts. *Journal of Pension Economics & Finance 13*(2), 172–209. Publisher: Cambridge University Press.

Goebel, J., M. M. Grabka, S. Liebig, M. Kroh, D. Richter, C. Schröder, and J. Schupp (2019). The German Socio-Economic Panel (SOEP). *Jahrbücher für Nationalökonomie und Statistik 239*(2), 345–360. Publisher: De Gruyter Oldenbourg Section: Jahrbücher für Nationalökonomie und Statistik.

Goldin, C. (2014). A Grand Gender Convergence: Its Last Chapter. *American Economic Review 104*(4), 1091–1119.

Grabka, M. M. and J. Goebel (2017). Realeinkommen sind von 1991 bis 2014 im Durchschnitt gestiegen: Erste Anzeichen für wieder zunehmende Einkommensungleichheit. *DIW-Wochenbericht 84*(4), 71–82. Publisher: Berlin: Deutsches Institut für Wirtschaftsforschung (DIW).

Grabka, M. M., J. Marcus, and E. Sierminska (2015). Wealth distribution within couples. *Review of Economics of the Household 13*(3), 459–486.

Guvenen, F., G. Kaplan, and J. Song (2014). The Glass Ceiling and The Paper Floor: Gender Differences among Top Earners, 1981-2012. Working Paper 20560, National Bureau of Economic Research. Series: Working Paper Series.

Guvenen, F., G. Kaplan, J. Song, and J. Weidner (2017). Lifetime incomes in the United States over six decades. Working Paper 23371, National Bureau of Economic Research. Series: Working Paper Series.

Haider, S. and G. Solon (2006). Life-cycle variation in the association between current and lifetime earnings. *American Economic Review 96*(4), 1308–1320.

Harkness, S. and J. Waldfogel (2003). The family gap in pay: evidence from seven industrialized countries. *Worker Well-Being and Public Policy 22*, 369–413.

Heien, T., K. Kortmann, and C. Schatz (2007). Altersvorsorge in Deutschland 2005, Technical report. Technical report, Deutsche Rentenversicherung Bund.

Hänisch, C. and J. Klos (2016). Long-run effects of career interruptions: A microsimulation study. Technical report, Discussion Paper Series, Wilfried Guth Endowed Chair for Constitutional Political Economy and Competition Policy.

Jann, B. (2008). The blinder-oaxaca decomposition for linear regression models. *Stata Journal 8*(4), 453–479(27).

Juhn, C. and K. McCue (2017). Specialization Then and Now: Marriage, Children, and the Gender Earnings Gap across Cohorts. *Journal of Economic Perspectives 31*(1), 183–204.

Kühhirt, M. and V. Ludwig (2012). Domestic work and the wage penalty for motherhood in West Germany. *Journal of Marriage and Family 74*(1), 186–200. Publisher: Wiley Online Library.

Killewald, A. and J. García-Manglano (2016). Tethered lives: A couple-based perspective on the consequences of parenthood for time use, occupation, and wages. *Social Science Research 60*, 266–282.

Killewald, A. and M. Gough (2013). Does Specialization Explain Marriage Penalties and Premiums? *American sociological review 78*(3), 477–502. Publisher: Sage Publications Sage CA: Los Angeles, CA.

Kleven, H. and C. Landais (2017). Gender inequality and economic development: fertility, education and norms. *Economica 84*(334), 180–209. Publisher: Wiley Online Library.

Kleven, H., C. Landais, J. Posch, A. Steinhauer, and J. Zweimüller (2019). The Impact of Family Policies on the Dynamics of Gender Inequality. *document de travail*.

Kleven, H., C. Landais, and J. E. Søgaard (2019). Children and Gender Inequality: Evidence from Denmark. *American Economic Journal: Applied Economics 11*(4), 181–209.

Levell, P. and J. Shaw (2015). Constructing full adult life-cycles from short panels. Technical report, IFS Working Papers.

Li, J. and C. O'Donoghue (2013). A survey of dynamic microsimulation models: uses, model structure and methodology. *International Journal of microsimulation 6*(2), 3–55. Publisher: International Microsimulation Association.

Neufeld, C. (2000). Alignment and variance reduction in DYNACAN. *Contributions to Economic Analysis 247*, 361–382. Publisher: NORTH HOLLAND.

Oaxaca, R. (1973). Male-female wage differentials in urban labor markets. *International economic review*, 693–709. Publisher: JSTOR.

Pence, K. M. (2006). The role of wealth transformations: An application to estimating the effect of tax incentives on saving. *The B.E. Journal of Economic Analysis & Policy 5*(1).

Plümper, T. and V. E. Troeger (2007). Efficient estimation of time-invariant and rarely changing variables in finite sample panel analyses with unit fixed effects. *Political Analysis 15*(2), 124–139. Publisher: Cambridge University Press.

Sierminska, E., D. Piazzalunga, and M. Grabka (2019). Transitioning towards more equality? wealth gender differences and the changing role of explanatory factors over time. SOEPpapers on Multidisciplinary Panel Data Research 1050, DIW Berlin, The German Socio-Economic Panel (SOEP).

Tyrowicz, J., L. van der Velde, and I. van Staveren (2018). Does Age Exacerbate the Gender-Wage Gap? New Method and Evidence From Germany, 1984–2014. *Feminist Economics 24*(4), 108–130. Publisher: Taylor & Francis.

Waldfogel, J. (1998). Understanding the" family gap" in pay for women with children. *Journal of economic Perspectives 12*(1), 137–156.

Westermeier, C., A. Rasner, and M. M. Grabka (2012). The prospects of the baby boomers: Methodological challenges in projecting the lives of an aging cohort. SOEPpapers on Multidisciplinary Panel Data Research 440, Berlin.

Zucchelli, E., A. M. Jones, and N. Rice (2010). The evaluation of health policies through microsimulation methods. *Health, Econometrics and Data Group (HEDG) Working Papers 10*(03).

Appendix A

A.1 Oaxaca Blinder Decomposition

The Oaxaca Blinder decomposition was simultaneously introduced by Oaxaca (1973) and Blinder (1973) and divides the gender differential in labor market outcomes (here: hourly wage or annual earnings) into an *endowment part* and a *coefficient part*. The endowment part of the gender differential accounts for the part of the gap which can be attributed to differences in the allocation of characteristics (e.g., working hours, highest level of education) between men and women. In contrast, the coefficient part captures the gender differences in labor market returns to characteristics, and therefore in their coefficients. In other words, it states the gender differences of what the labor market is willing to pay for the same characteristics. This part is also called the raw or adjusted gender wage/earnings differential. This adjusted gap, however, also contains the effects of gender differences in unobserved predictors (Jahn 2008). The Oaxaca-Blinder decomposition approach enables us to analyze whether the gender gap in wages/earnings is mainly driven by the different distributions of productivity characteristics or by different rewards for these characteristics by gender.

The gender gap G_x is defined as the difference between the means of the labor market outcomes L at age x of men m and women f:

$$G_x = E(L_{mx}) - E(L_{fx}) \tag{A.1}$$

L_{sx} for either sex (s) is based on the linear model

$$L_{sx} = Z'_{sx}\beta_{sx} + \epsilon_{sx}, \quad E(\epsilon_{sx}) = 0, \tag{A.2}$$

where the vector Z includes all relevant characteristics, β is the estimation vector and ϵ is the error term. Inserting Equation (A.2) into Equation (A.1), the earnings differential can also be written as:

$$G_x = E(Z_{mx})'\beta_{mx} - E(Z_{fx})'\beta_{fx}. \tag{A.3}$$

For the decomposition of the results, a non-discriminatory coefficient vector is needed, called β^*. Following Neumark (1988), the vector is determined as a pooled regression over both sexes. The gender gap can then be rewritten as:

$$G_x = \underbrace{[E(Z_{mx}) - E(Z_{fx})]'\beta_x^*}_{\text{Endowment part}} + \underbrace{[E(Z_{mx})'(\beta_{mx} - \beta_x^*) + E(Z_{fx})'(\beta_x^* - \beta_{fx})]}_{\text{Coefficient part}} \tag{A.4}$$

where the first part of equation (A.4) is the endowment part and the second part is the coefficient component of the gender gap in the labor market outcome.

A.2 Supplementary Results: Oaxaca Blinder Decomposition

Table A.1: Regression Results for Hourly Wages - Women

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Age	20	25	30	35	40	45	50	55	60
One child	-0.019	-0.061	-0.159***	-0.023	-0.003	0.075**	0.068*	0.104**	-0.058
	(0.158)	(0.055)	(0.046)	(0.036)	(0.038)	(0.034)	(0.040)	(0.048)	(0.070)
Two children	-0.461	-0.234**	-0.154***	-0.058	-0.000	0.073**	0.138***	0.088*	-0.014
	(0.501)	(0.107)	(0.058)	(0.041)	(0.039)	(0.036)	(0.040)	(0.048)	(0.071)
3 or more children		-0.171	-0.111	-0.167***	-0.030	0.036	0.129***	0.103*	-0.051
		(0.195)	(0.093)	(0.055)	(0.050)	(0.043)	(0.048)	(0.058)	(0.083)
Married	0.033	-0.038	0.054	0.008	0.004	0.068**	-0.052*	-0.031	0.055
	(0.100)	(0.036)	(0.034)	(0.030)	(0.029)	(0.027)	(0.029)	(0.035)	(0.048)
Years FT	0.445***	0.055**	0.059***	0.028***	0.035***	0.026***	0.027***	0.014**	0.030***
	(0.061)	(0.022)	(0.014)	(0.009)	(0.007)	(0.005)	(0.005)	(0.006)	(0.007)
Years FT (sq)	-0.054***	-0.003	-0.002***	0.000	-0.001*	-0.000	-0.000	-0.000	-0.000**
	(0.016)	(0.002)	(0.001)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Years PT	-0.023	-0.039	-0.028	-0.019*	-0.013	-0.020***	-0.002	-0.001	-0.011
	(0.103)	(0.029)	(0.017)	(0.011)	(0.008)	(0.006)	(0.006)	(0.006)	(0.008)
Years PT (sq)	0.012	0.001	0.003*	0.002**	0.001**	0.001***	0.000*	0.000	0.000**
	(0.028)	(0.004)	(0.002)	(0.001)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Years UE	-0.689***	-0.101**	-0.174***	0.013	-0.062***	-0.096***	-0.076***	-0.058***	-0.044*
	(0.209)	(0.043)	(0.039)	(0.021)	(0.018)	(0.018)	(0.018)	(0.017)	(0.024)
Years UE (sq)	0.236*	-0.000	0.034***	-0.005**	0.004**	0.007***	0.004**	0.001	0.001
	(0.140)	(0.006)	(0.009)	(0.003)	(0.002)	(0.002)	(0.002)	(0.001)	(0.002)
Part-time	0.244***	0.272***	0.110***	0.182***	0.161***	0.102***	0.023	-0.068*	0.069
	(0.081)	(0.049)	(0.042)	(0.032)	(0.031)	(0.027)	(0.031)	(0.039)	(0.054)
Education	-0.068***	0.007	-0.033**	-0.063***	0.018	0.008	-0.023	0.012	-0.052
	(0.024)	(0.017)	(0.015)	(0.015)	(0.023)	(0.015)	(0.021)	(0.029)	(0.036)
Education (sq)	0.004*	0.001	0.003***	0.006***	0.002***	0.003***	0.004***	0.002**	0.005***
	(0.002)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)
Constant	2.241***	1.714***	1.233***	1.700***	1.825***	1.508***	1.033***	1.614***	1.898***
	(0.770)	(0.266)	(0.211)	(0.222)	(0.199)	(0.161)	(0.313)	(0.373)	(0.463)
Obs.	382	882	1307	1859	2493	2653	2043	1320	778
R-squared	0.323	0.127	0.187	0.240	0.192	0.219	0.205	0.213	0.248
Cohort-FE	YES	YES	YES	YES	YES	YES	YES	YES	YES
Sector-FE	YES	YES	YES	YES	YES	YES	YES	YES	YES

Note: Standard errors in parentheses; The stars refer to the following significance level: $*p < 0.1$, $**p < 0.05$, $***p < 0.01$.
Source: Own calculations based on SOEP v.35 (1984-2018).

Table A.2: Regression Results for Hourly Wages - Men

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Age	20	25	30	35	40	45	50	55	60
One child	0.160	0.089	-0.017	-0.017	0.007	0.029	-0.043	0.026	0.053
	(0.349)	(0.055)	(0.031)	(0.024)	(0.022)	(0.024)	(0.029)	(0.040)	(0.053)
Two children	-0.952	0.134*	0.065*	0.069***	0.033	0.103***	-0.000	0.034	-0.010
	(0.751)	(0.079)	(0.035)	(0.024)	(0.022)	(0.023)	(0.027)	(0.037)	(0.049)
3 or more children	-0.006	0.139*	-0.024	0.015	0.050*	0.049*	-0.013	0.075	0.203***
	(0.173)	(0.075)	(0.045)	(0.033)	(0.028)	(0.029)	(0.035)	(0.049)	(0.067)
Married	-0.026	0.015	0.164***	0.085***	0.101***	0.069***	0.084***	0.020	0.105**
	(0.166)	(0.041)	(0.027)	(0.023)	(0.021)	(0.023)	(0.026)	(0.035)	(0.048)
Years FT	0.737***	0.164***	0.105***	0.058***	0.053***	0.045***	0.032***	0.040***	-0.030
	(0.061)	(0.023)	(0.013)	(0.008)	(0.007)	(0.008)	(0.009)	(0.012)	(0.036)
Years FT (sq)	-0.108***	-0.012***	-0.006***	-0.002***	-0.001***	-0.001***	-0.001***	-0.001**	0.001
	(0.015)	(0.002)	(0.001)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.001)
Years PT	0.225	-0.197***	-0.074***	-0.020*	-0.056***	-0.038***	-0.069***	-0.070***	-0.057***
	(0.181)	(0.036)	(0.018)	(0.012)	(0.010)	(0.011)	(0.010)	(0.014)	(0.017)
Years PT (sq)	-0.079	0.021***	0.006***	0.001	0.003***	0.003***	0.003***	0.003***	0.001*
	(0.071)	(0.005)	(0.002)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)
Years UE	-0.279	-0.108**	-0.178***	-0.113***	-0.117***	-0.105***	-0.094***	-0.069***	-0.094***
	(0.195)	(0.050)	(0.029)	(0.015)	(0.013)	(0.012)	(0.014)	(0.019)	(0.036)
Years UE (sq)	0.099	-0.005	0.023***	0.004***	0.006***	0.006***	0.004***	0.002**	0.006
	(0.101)	(0.014)	(0.006)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.004)
Part-time	0.271***	0.422***	0.257***	0.173***	0.336***	0.251***	0.189***	0.389***	0.201***
	(0.075)	(0.054)	(0.043)	(0.034)	(0.031)	(0.031)	(0.039)	(0.045)	(0.054)
Education	-0.031	-0.038***	-0.071***	-0.039***	-0.051***	0.040*	0.060***	0.002	0.020
	(0.030)	(0.013)	(0.010)	(0.010)	(0.013)	(0.022)	(0.023)	(0.055)	(0.067)
Education (sq)	0.003	0.003***	0.005***	0.003***	0.004***	0.001*	0.000	0.002	0.002
	(0.003)	(0.001)	(0.000)	(0.000)	(0.000)	(0.001)	(0.001)	(0.002)	(0.002)
Constant	1.996***	1.591***	2.263***	2.331***	2.519***	1.703***	1.673***	1.875***	1.318
	(0.504)	(0.163)	(0.183)	(0.172)	(0.173)	(0.187)	(0.231)	(0.437)	(0.926)
Obs.	383	900	1746	2464	2863	2939	2261	1546	980
R-squared	0.449	0.231	0.185	0.229	0.277	0.283	0.252	0.184	0.208
Cohort-FE	YES	YES	YES	YES	YES	YES	YES	YES	YES
Sector-FE	YES	YES	YES	YES	YES	YES	YES	YES	YES

Note: Standard errors in parentheses; The stars refer to the following significance level: $*p < 0.1$, $**p < 0.05$, $***p < 0.01$.
Source: Own calculations based on SOEP v.35 (1984-2018).

Table A.3: Regression Results for Annual Earnings - Women

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Age	20	25	30	35	40	45	50	55	60
One child	0.160	0.089	-0.017	-0.017	0.007	0.029	-0.043	0.026	0.053
	(0.349)	(0.055)	(0.031)	(0.024)	(0.022)	(0.024)	(0.029)	(0.040)	(0.053)
Two children	-0.952	0.134*	0.065*	0.069***	0.033	0.103***	-0.000	0.034	-0.010
	(0.751)	(0.079)	(0.035)	(0.024)	(0.022)	(0.023)	(0.027)	(0.037)	(0.049)
3 or more children	-0.006	0.139*	-0.024	0.015	0.050*	0.049*	-0.013	0.075	0.203***
	(0.173)	(0.075)	(0.045)	(0.033)	(0.028)	(0.029)	(0.035)	(0.049)	(0.067)
Married	-0.026	0.015	0.164***	0.085***	0.101***	0.069***	0.084***	0.020	0.105**
	(0.166)	(0.041)	(0.027)	(0.023)	(0.021)	(0.023)	(0.026)	(0.035)	(0.048)
Years FT	0.737***	0.164***	0.105***	0.058***	0.053***	0.045***	0.032***	0.040***	-0.030
	(0.061)	(0.023)	(0.013)	(0.008)	(0.007)	(0.008)	(0.009)	(0.012)	(0.036)
Years FT (sq)	-0.108***	-0.012***	-0.006***	-0.002***	-0.001***	-0.001***	-0.001***	-0.001**	0.001
	(0.015)	(0.002)	(0.001)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.001)
Years PT	0.225	-0.197***	-0.074***	-0.020*	-0.056***	-0.038***	-0.069***	-0.070***	-0.057***
	(0.181)	(0.036)	(0.018)	(0.012)	(0.010)	(0.011)	(0.010)	(0.014)	(0.017)
Years PT (sq)	-0.079	0.021***	0.006***	0.001	0.003***	0.003***	0.003***	0.003***	0.001*
	(0.071)	(0.005)	(0.002)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)
Years UE	-0.279	-0.108**	-0.178***	-0.113***	-0.117***	-0.105***	-0.094***	-0.069***	-0.094***
	(0.195)	(0.050)	(0.029)	(0.015)	(0.013)	(0.012)	(0.014)	(0.019)	(0.036)
Years UE (sq)	0.099	-0.005	0.023***	0.004***	0.006***	0.006***	0.004***	0.002**	0.006
	(0.101)	(0.014)	(0.006)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.004)
Part-time	0.271***	0.422***	0.257***	0.173***	0.336***	0.251***	0.189***	0.389***	0.201***
	(0.075)	(0.054)	(0.043)	(0.034)	(0.031)	(0.031)	(0.039)	(0.045)	(0.054)
Education	-0.031	-0.038***	-0.071***	-0.039***	-0.051***	0.040*	0.060***	0.002	0.020
	(0.030)	(0.013)	(0.010)	(0.010)	(0.013)	(0.022)	(0.023)	(0.055)	(0.067)
Education (sq)	0.003	0.003***	0.005***	0.003***	0.004***	0.001*	0.000	0.002	0.002
	(0.003)	(0.001)	(0.000)	(0.000)	(0.000)	(0.001)	(0.001)	(0.002)	(0.002)
Constant	1.996***	1.591***	2.263***	2.331***	2.519***	1.703***	1.673***	1.875***	1.318
	(0.504)	(0.163)	(0.183)	(0.172)	(0.173)	(0.187)	(0.231)	(0.437)	(0.926)
Obs.	383	900	1746	2464	2863	2939	2261	1546	980
R-squared	0.449	0.231	0.185	0.229	0.277	0.283	0.252	0.184	0.208
Cohort-FE	YES	YES	YES	YES	YES	YES	YES	YES	YES
Sector-FE	YES	YES	YES	YES	YES	YES	YES	YES	YES

Note: Standard errors in parentheses; The stars refer to the following significance level: $*p < 0.1$, $**p < 0.05$, $***p < 0.01$.
Source: Own calculations based on SOEP v.35 (1984-2018).

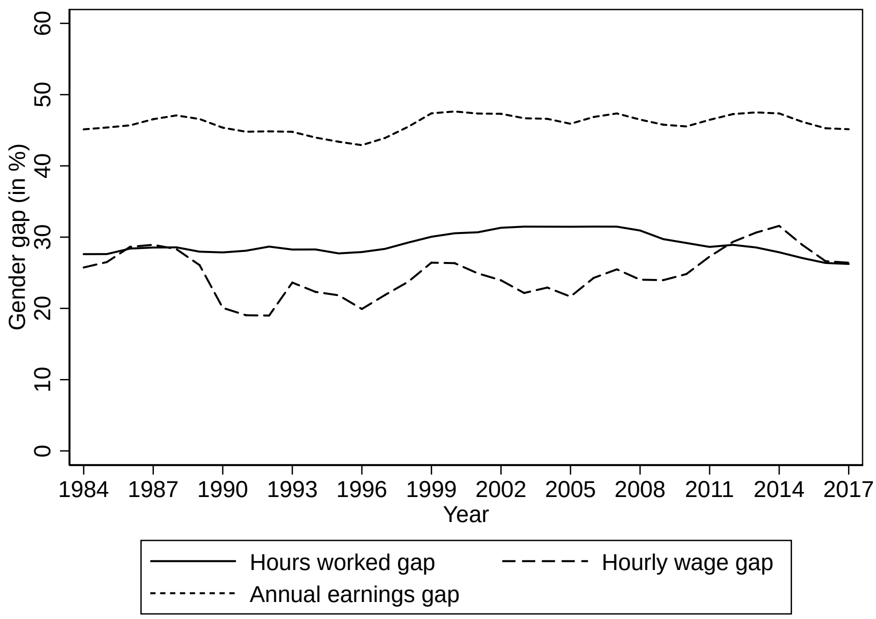
Table A.4: Regression Results for Annual Earnings - Men

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Age	20	25	30	35	40	45	50	55	60
One child	0.189	0.061	-0.008	-0.014	0.014	0.028	-0.042	-0.003	0.073
	(0.332)	(0.052)	(0.028)	(0.023)	(0.021)	(0.024)	(0.028)	(0.038)	(0.051)
Two children	-1.006	0.139*	0.066**	0.067***	0.026	0.091***	-0.008	0.019	0.015
	(0.714)	(0.075)	(0.032)	(0.023)	(0.021)	(0.022)	(0.026)	(0.036)	(0.047)
3 or more children	-0.008	0.128*	-0.026	0.025	0.050*	0.051*	0.013	0.100**	0.157**
	(0.164)	(0.071)	(0.041)	(0.032)	(0.026)	(0.028)	(0.034)	(0.047)	(0.065)
Married	0.017	0.021	0.121***	0.078***	0.089***	0.054**	0.091***	0.028	0.099**
	(0.158)	(0.039)	(0.025)	(0.022)	(0.020)	(0.022)	(0.025)	(0.034)	(0.046)
Years FT	0.731***	0.179***	0.094***	0.058***	0.045***	0.043***	0.031***	0.052***	-0.033
	(0.058)	(0.022)	(0.012)	(0.008)	(0.006)	(0.008)	(0.009)	(0.012)	(0.035)
Years FT (sq)	-0.106***	-0.013***	-0.005***	-0.002***	-0.001***	-0.001***	-0.001**	-0.001***	0.001
	(0.014)	(0.002)	(0.001)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Years PT	0.095	-0.191***	-0.044***	-0.020*	-0.060***	-0.042***	-0.057***	-0.079***	-0.056***
	(0.174)	(0.034)	(0.016)	(0.011)	(0.010)	(0.011)	(0.010)	(0.013)	(0.015)
Years PT (sq)	-0.051	0.024***	0.003	0.000	0.003***	0.003***	0.003***	0.003***	0.001*
	(0.067)	(0.005)	(0.002)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)
Years UE	-0.204	-0.107**	-0.188***	-0.110***	-0.118***	-0.101***	-0.091***	-0.109***	-0.122***
	(0.186)	(0.047)	(0.026)	(0.014)	(0.012)	(0.012)	(0.014)	(0.018)	(0.035)
Years UE (sq)	0.040	-0.005	0.023***	0.004***	0.006***	0.005***	0.004***	0.004***	0.008**
	(0.097)	(0.014)	(0.005)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.004)
Weekly hours	0.044***	0.049***	0.070***	0.052***	0.037***	0.054***	0.077***	0.053***	0.071***
	(0.009)	(0.004)	(0.003)	(0.003)	(0.003)	(0.003)	(0.004)	(0.004)	(0.005)
Weekly hours (sq)	-0.000***	-0.000***	-0.001***	-0.000***	-0.000***	-0.000***	-0.001***	-0.000***	-0.001***
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Education	-0.018	-0.040***	-0.076***	-0.041***	-0.057***	0.039*	0.053**	0.091*	0.019
	(0.028)	(0.012)	(0.009)	(0.009)	(0.012)	(0.021)	(0.023)	(0.053)	(0.065)
Education (sq)	0.003	0.004***	0.005***	0.004***	0.004***	0.002**	0.001	-0.001	0.002
	(0.002)	(0.001)	(0.000)	(0.000)	(0.000)	(0.001)	(0.001)	(0.002)	(0.002)
Constant	7.710***	8.367***	8.434***	8.715***	9.163***	7.965***	7.506***	7.527***	8.427***
	(0.195)	(0.113)	(0.099)	(0.103)	(0.115)	(0.174)	(0.201)	(0.394)	(0.730)
Obs.	383	900	1746	2464	2863	2939	2261	1546	980
R-squared	0.542	0.539	0.481	0.394	0.409	0.417	0.437	0.400	0.522
Cohort-FE	YES	YES	YES	YES	YES	YES	YES	YES	YES
Sector-FE	YES	YES	YES	YES	YES	YES	YES	YES	YES

Note: Standard errors in parentheses; The stars refer to the following significance level: $*p < 0.1$, $**p < 0.05$, $***p < 0.01$.
Source: Own calculations based on SOEP v.35 (1984-2018).

A.3 Supplementary Results: Gender Gaps

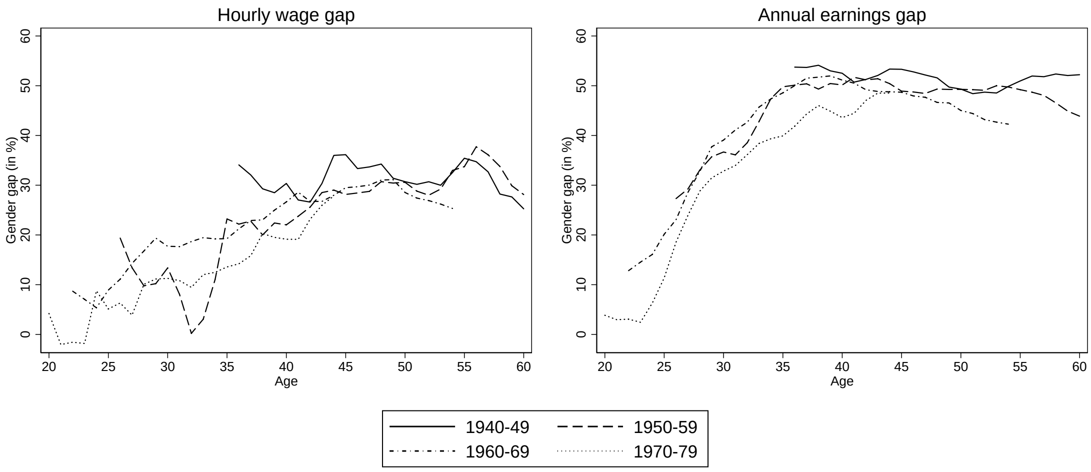
Figure A.1: Gender Gaps in Labor Market Outcomes by Survey Year



Notes: Only employed individuals are considered. Does not include values of zero annual earnings. Cohorts 1940-1979, weighted sample. Annual earnings and hourly wages in Euro.

Source: Own calculations based on SOEP v.35 (1984-2018).

Figure A.2: Gender Gaps in Hourly Wages and Annual Earnings by Cohort

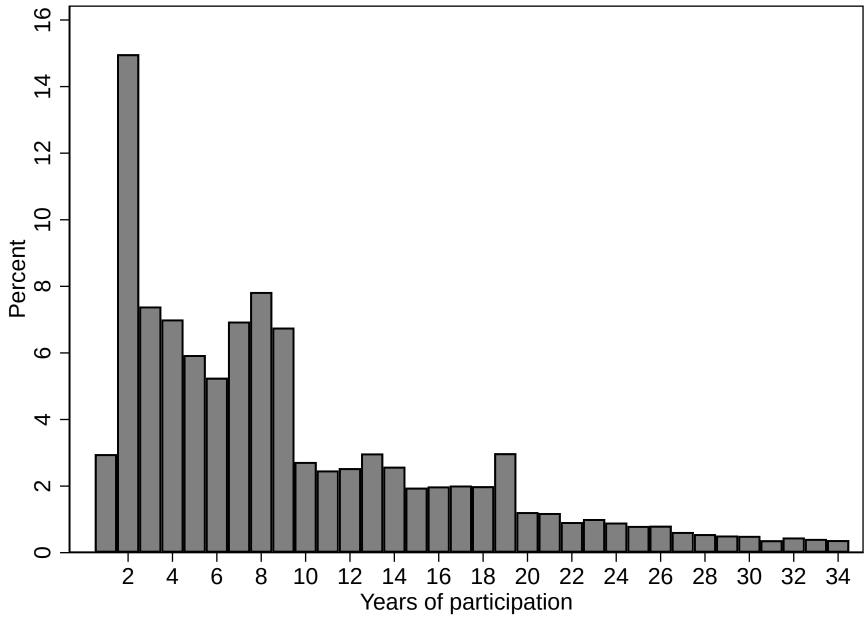


Notes: Only employed individuals are considered. Does not include values of zero hourly wages or annual earnings. Cohorts 1940-1979, weighted sample.

Source: Own calculations based on SOEP v.35 (1984-2018).

Appendix B. Estimation results for full time employees and without civil servants

Figure B.1: Distribution of Participation Years in the SOEP



Notes: Refers to participation years of the SOEP sample used in Chapter 2 of this paper. Restrictions for the microsimulation in Chapter 3 are not applied here.

Source: Own calculations based on SOEP v.33 (1991-2015).

Table B.1: Distribution of Cohorts by Gender

Birth Cohort	Number of Men	Number of Women	Total
1964	382	324	706
1965	373	383	756
1966	404	425	829
1967	385	401	786
1968	378	385	763
1969	388	387	775
1970	311	364	675
1971	303	342	645
1972	288	304	592
Total	3212	3315	6527

Source: Own calculations based on SOEP v.35 (1984-2018).

Table B.2: Overview Regression Models of the Dynamic Microsimulation

Dependent Variables (Model)	Explanatory Variables and Subsamples
Child birth in $t+1$ (*Logit*)	Number of children, age of youngest child, earnings;Additionally, for married women: partner's age, highest level of education and earnings; Run separately for married women and single women
Change in marital status (married/single) in $t+1$ (*Logit*)	Marriage duration term interacted with age, number of children; Additionally, for women: age of youngest child; Additionally, for married individuals: Partner's age and highest level of education; Run separately for men and women for each respective marital status
Change in labor force status in t+1: (*Logit*)	Labor force status in t and $t-1$, labor market history (years in full-time, part-time, unemployment), number of children (not for unmarried men); Additionally, for women: number of years since birth of last child; Additionally, for married individuals: partner's labor force status and earnings in t; Run separately for men and women for each respective marital status
Change in employment status (working/ unemployed) in $t+1$ (*Logit*)	Employment status in $t-1$, labor market history (years in full-time, part-time, unemployment), number of children (not for unmarried men); Additionally, for women: number of years since birth of last child; Additionally, for married individuals: partner's employment status and earnings of the in t;Run separately for men and women for each possible combination of marital and employment status in t
Transition in employment or unemployment in t+1 after not participating in the labor market in t (*Logit*)	Employment status in $t-1$, labor market history (years in full-time, part-time, unemployment), number of children (not for unmarried men); Additionally, for women: number of years since birth of last child-, Additionally, for married individuals: partner's employment status and earnings in t; Run separately for men and women for each respective marital status (requirement: participating in the labor market in $t+1$)
Transition full-time work/ part-time work in $t+1$ (*Logit*)	Labor force status in $t-1$, dummy variable indicating full-time or part-time work in $t-1$, labor market history (years in full-time, part-time, unemployment), number of children (not for unmarried men); Additionally, for women: number of years since birth of last child; Additionally, for married individuals: partner's employment status and earnings of the partner in t; Run separately for men and women for each possible combination of marital and full-time/ part-time status in t
Transition in full-time work/ part-time work in $t+1$ after not working in t (*Logit*)	Labor force status in $t-1$, dummy variable indicating full-time or part-time work in $t-1$, labor market history (years in full-time, part-time, unemployment), number of children (not for unmarried men)-, Additionally, for women: number of years since birth of last child; Additionally, for married individuals: employment status and earnings of the partner in t; Run separately for men and women for each respective marital status (requirement: working in $t+1$)
Number of working hours in t (*OLS*)	Annual hours worked in $t-1$ and $t-2$, annual earnings in $t-1$, dummy variable indicating full-time or part-time work in and labor market status $t-1$, number of children (not for unmarried men); Additionally, for married individuals: earnings of the partner in t-1; Run separately for men and women for each respective marital and work (full-time/part-time) status
Annual earnings in t (*OLS*)	Annual earnings in $t-1$ and $t-2$, annual hours worked in t, $t-1$ and $t-2$, labor market history (years in full-time, part-time, unemployment), dummy indicating marital status; Run separately for men and women

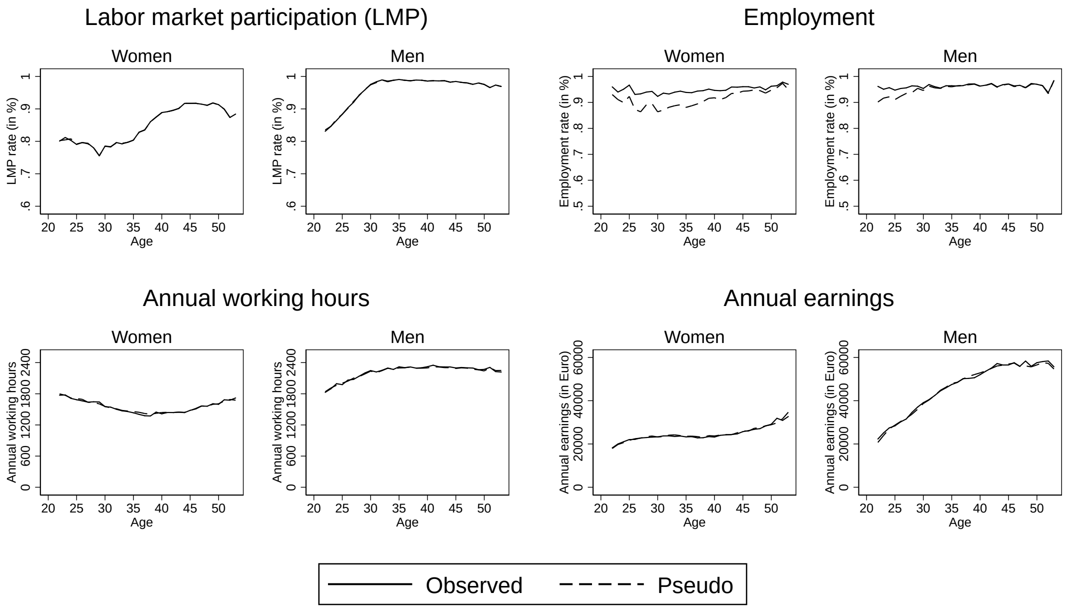
Notes: Explanatory variables which are included in every model: highest level of education and year of birth interacted with (quadratic) age, place of residency before 1989 (East or West Germany), immigration background (yes or no).
Source: Own calculations based on SOEP v.35 (1984-2018).

B.1 Robustness: Microsimulation

B.1.1 Pseudo Missings

To test the robustness of our simulation model further, we use the concept of pseudo missings. To that end, we set truly observed years for some part of the sample missing (pseudo missings) and predict their now missing observations again by using our dynamic microsimulation and the regression coefficients previously obtained. As we need a starting point of at least two observations for our models due to the lagged terms, we use the first two truly observed years for everyone before starting to create pseudo missings. Figure B.2 shows the differences between the simulated pseudo missings (dashed line) and the truly observed information (solid line) for labor force status, employment status, annual working hours and annual earnings. In most graphs, the level of accuracy of the model is so high that it is hard to even tell the solid and dashed line apart. For labor market status, the model predicts 99.9 percent of all pseudo missings correctly. And even for employment status, where there appear to be bigger differences between pseudo missing and observations at a first glance, overall 97.7 percent of all cases are simulated correctly. These results further support the robustness of our simulation model.

Figure B.2: Pseudo Missings for Labor Market Outcomes



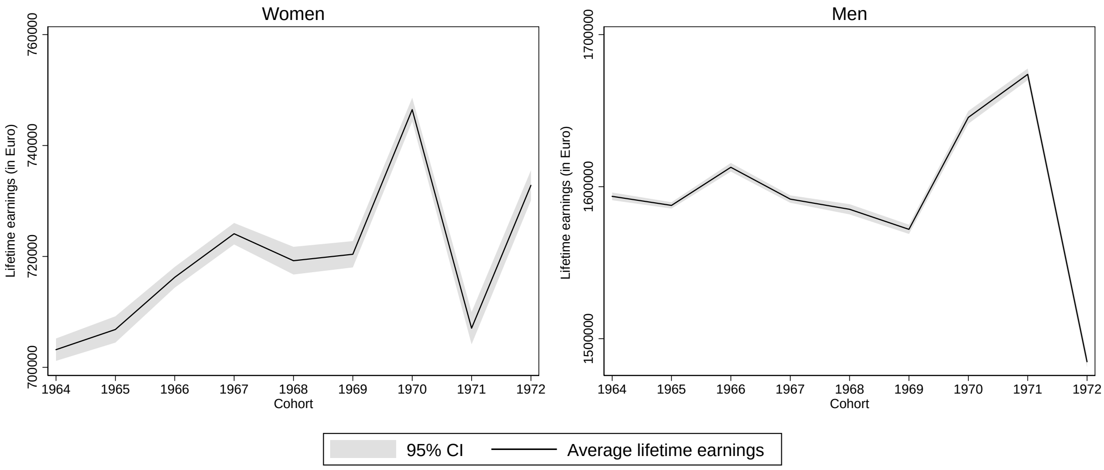
Notes: The graphs comparing truly observed and simulated pseudo information for annual working hours and annual earnings only focus on employed individuals.

Source: Own calculations based on SOEP v.35 (1984-2018).

B.1.2 Monte Carlo Simulation

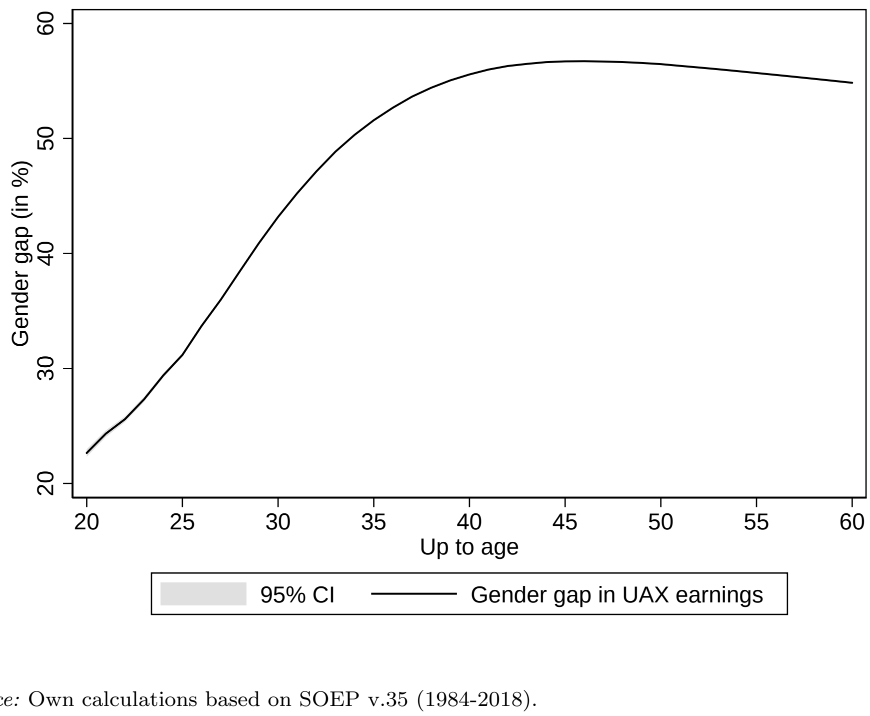
Another way to validate the robustness of our dynamic microsimulation model is to make use of the underlying random process described in Section 3.1.1. We implement a Monte Carlo simulation approach by simulating each individual's employment biographies 100 times. By doing so, due to the underlying random process determining transitions in labor market outcome variables between $t-1$ and t, we simulate up to 100 different employment biographies for each individual. However, due to limited computational capacities we only simulate the employment variables (labor market status, employment status, full-time/part-time work, annual working hours and annual earnings) and keep the family information (number of children and marital status) constant for each of the 100 iterations. In the next step, we calculate lifetime earnings for each of the 100 simulated career paths per individual and compute the average lifetime earnings and the resulting UAX earnings gender gap in the population for each of the 100 runs. By deriving the 95% confidence intervals we can analyze whether average lifetime earnings vary significantly for different underlying random processes or whether they are robust. The results are presented in Figures B.3 and B.4. Figure B.3 shows that lifetime earnings by cohorts are very robust. However, lifetime earnings of women vary more strongly than men's do. Figure B.4 provides evidence for a very narrow 95% confidence interval for the gender gap in UAX earnings. Consequently, the results of the Monte Carlo simulation confirm the high robustness of our simulation outcomes.

Figure B.3: Monte Carlo Simulation for Earnings



Source: Own calculations based on SOEP v.35 (1984-2018).

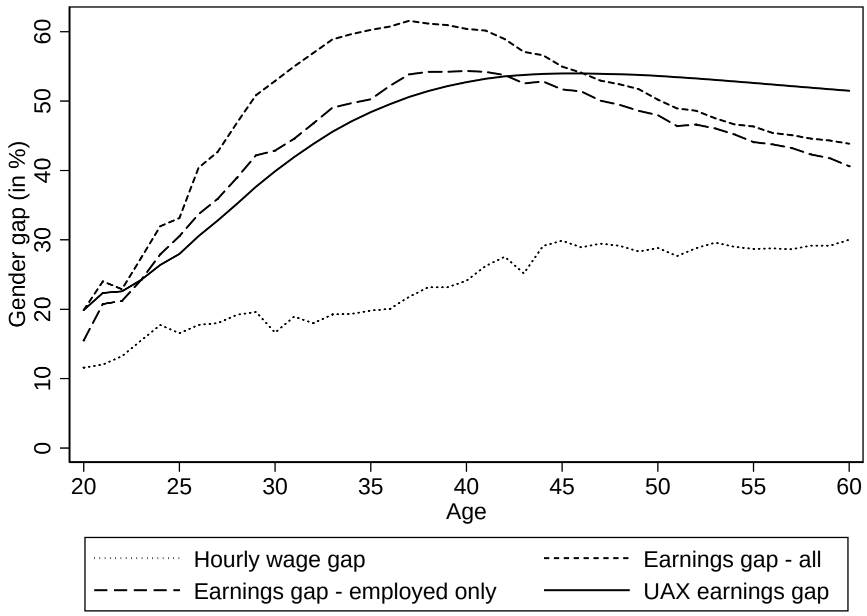
Figure B.4: Monte Carlo Simulation for the Gender Gap in UAX Earnings



Source: Own calculations based on SOEP v.35 (1984-2018).

Appendix C: Supplementary Material

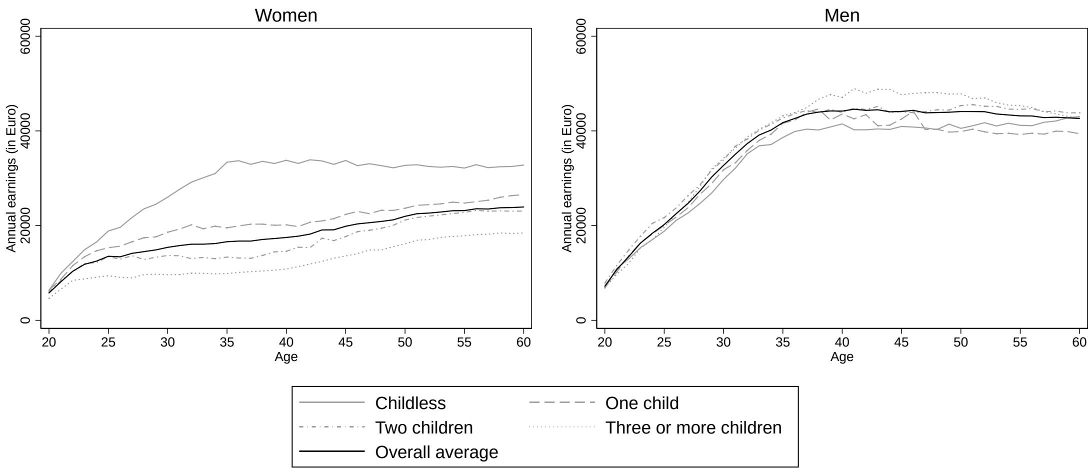
Figure C.1: Gender Gaps in Earnings by Different Concepts



Notes: Individuals with zero UAX earnings are included in the calculation. For annual earnings gap, all employed and unemployed individuals are considered. Cohorts 1964-1970.

Source: Own calculations based on SOEP v.35 (1984-2018).

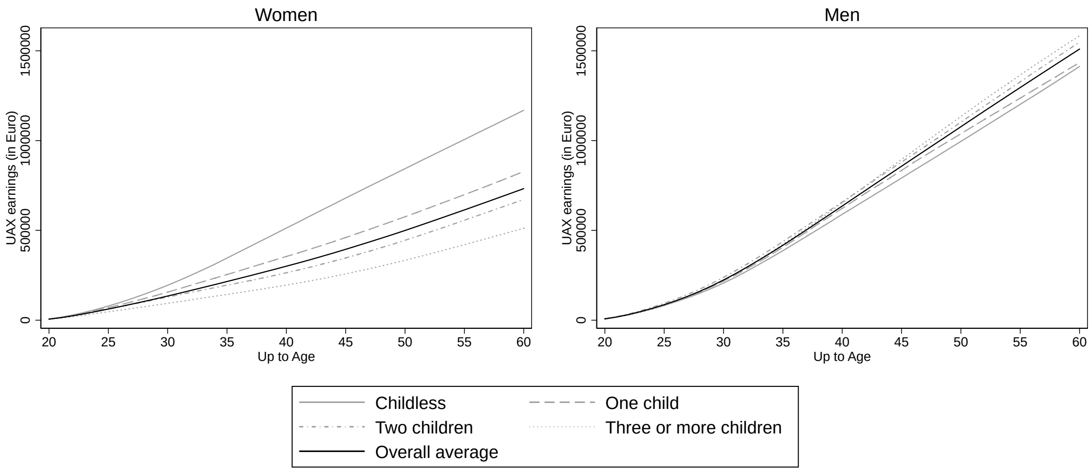
Figure C.2: Annual Earnings by Gender and Number of Children



Notes: Employed and unemployed individuals are considered. Number of children refers to the total number at age 50. Cohorts 1964-1972.

Source: Own calculations based on SOEP v.35 (1984-2018).

Figure C.3: UAX Earnings by Gender and Number of Children



Notes: Employed and unemployed individuals are considered. Number of children refers to the total number at age 50. Cohorts 1964-1972.

Source: Own calculations based on SOEP v.35 (1984-2018).

Diskussionsbeiträge - Fachbereich Wirtschaftswissenschaft - Freie Universität Berlin
Discussion Paper - School of Business & Economics - Freie Universität Berlin

2022 erschienen:

2022/1 AHRENS, Steffen; Ciril BOSCH-ROSA und Thomas MEISSNER:
Intertemporal Consumption and Debt Aversion: A Replication and Extension
Economics

2022/2 WOLF, Elias: Estimating Growth at Risk with Skewed Stochastic Volatility Models
Economics